JGR Oceans

RESEARCH ARTICLE
10.1029/2020JC017060

Special Section:
Advancing prediction of coastal marine ecosystems

Seasonal Forecasting Skill of Sea-Level Anomalies in a Multi-Model Prediction Framework

Xiaoyu Long[1], Matthew J. Widlansky[1], Claire M. Spillman[2], Arun Kumar[3], Magdalena Balmaseda[4], Philip R. Thompson[1,5], Yoshimitsu Chikamoto[6], Grant A. Smith[2], Bohua Huang[7], Chul-Su Shin[7], Mark A. Merrifield[8], William V. Sweet[9], Eric Leuliette[10], H. S. Annamalai[5,11], John J. Marra[12], and Gary Mitchum[13]

[1]Joint Institute for Marine and Atmospheric Research, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Honolulu, HI, USA, [2]Bureau of Meteorology, Melbourne, VIC, Australia, [3]Climate Prediction Center, NCEP/NWS/NOAA, College Park, MD, USA, [4]European Centre for Medium-Range Weather Forecasts, Reading, UK, [5]Department of Oceanography, University of Hawai'i at Mānoa, Honolulu, HI, USA, [6]Department of Plants, Soils and Climate, Utah State University, Logan, UT, USA, [7]Department of Atmospheric, Oceanic, and Earth Sciences, College of Science, George Mason University, Fairfax, VA, USA, [8]Scripps Institution of Oceanography, University of California, San Diego, CA, USA, [9]NOAA/National Ocean Service, Silver Spring, MD, USA, [10]NOAA/NWS NCWCP Laboratory for Satellite Altimetry, College Park, MD, USA, [11]International Pacific Research Center, University of Hawai'i at Mānoa, Honolulu, HI, USA, [12]NOAA/NESDIS/National Centers for Environmental Information, Inouye Regional Center, Honolulu, HI, USA, [13]College of Marine Science, University of South Florida, St. Petersburg, FL, USA

Key Points:

- Prediction skill of seasonal sea-level anomalies up to a year in the future is assessed in 10 global climate forecasting systems
- Skillful seasonal sea-level forecasts are found in the tropics, whereas the skill is lower in higher latitudes and along continental coasts
- The most skillful predictions are from models with more accurate initializations of sea level and higher resolutions of the ocean

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Long,
xlong@hawaii.edu

Citation:

Long, X., Widlansky, M. J., Spillman, C. M., Kumar, A., Balmaseda, M., Thompson, P. R., et al. (2021). Seasonal forecasting skill of sea-level anomalies in a multi-model prediction framework. *Journal of Geophysical Research: Oceans*, *126*, e2020JC017060. https://doi.org/10.1029/2020JC017060

Received 11 DEC 2020
Accepted 18 MAY 2021

Abstract Coastal high water level events are increasing in frequency and severity as global sea-levels rise, and are exposing coastlines to risks of flooding. Yet, operational seasonal forecasts of sea-level anomalies are not made for most coastal regions. Advancements in forecasting climate variability using coupled ocean-atmosphere global models provide the opportunity to predict the likelihood of future high water events several months in advance. However, the skill of these models to forecast seasonal sea-level anomalies has not been fully assessed, especially in a multi-model framework. Here, we construct a 10-model ensemble of retrospective forecasts with future lead times of up to 11 months. We compare predicted sea levels from bias-corrected forecasts with 20 years of observations from satellite-based altimetry and shore-based tide gauges. Forecast skill, as measured by anomaly correlation, tends to be highest in the tropical and subtropical open oceans, whereas the skill is lower in the higher latitudes and along some continental coasts. For most locations, multi-model averaging produces forecast skill that is comparable to or better than the best performing individual model. We find that the most skillful predictions typically come from forecast systems with more accurate initializations of sea level, which is generally achieved by assimilating altimetry data. Having relatively higher horizontal resolution in the ocean is also beneficial, as such models seem to better capture dynamical processes necessary for successful forecasts. The multi-model assessment suggests that skillful seasonal sea-level forecasts are possible in many, though not all, parts of the global ocean.

Plain Language Summary We assess 10 global climate forecasting systems to predict monthly and seasonal anomalies of local sea levels up to a year into the future. We find that skillful seasonal sea-level forecasts are possible in many parts of the global ocean. Forecast skill is generally highest in the tropical and subtropical open oceans, whereas the skill is lower in the higher latitudes and along continental coasts. For most locations, multi-model averaging improves the forecast skill, compared to considering the models individually. Overall, the most skillful predictions are from forecasting systems with more accurate initializations of sea level and higher horizontal resolutions of the ocean.

1. Introduction

Sea-level variability increasingly contributes to coastal impacts such as flooding, erosion, and damage to infrastructure or ecosystems due to saltwater inundation (Anderson et al., 2010; Becker et al., 2014; Cazenave & Cozannet, 2014; Moftakhari et al., 2015; Nicholls & Cazenave, 2010). The longer-term increased occurrence of these impacts is mostly associated with global sea-level rise that is ongoing (Church & White, 2006; Fasullo & Nerem, 2018), and is projected to accelerate with continued greenhouse warming (Church et al., 2013; Sweet & Park, 2014; Yin, 2012). However, shorter-term and more localized variations

in the sea level explain much of the monthly-to-seasonal uncertainty of when coastal impacts occur (Calafat et al., 2018; Hamlington et al., 2015; Long et al., 2020; Merrifield et al., 2013; Palanisamy et al., 2014; Piecuch et al., 2016; Sweet & Zervas, 2011). In contrast, astronomical tidal cycles provide a useful deterministic constraint on when coastal water levels are likely to exceed the average high tide (e.g., clustering of coastal impacts around a so-called King Tide event; Román-Rivera & Ellis, 2018; Thompson et al., 2019). Vertical land motions provide further constraints on changes in the likelihood for coastal impacts related to sea level (i.e., a subsiding location will become increasing susceptible to high tides, as well as above-normal sea-level anomalies; Wöppelmann & Marcos, 2016). Storm surges, wave setups, tsunamis, and other processes that contribute to sub-monthly sea-level variability, while not directly considered here, have coastal impacts that can potentially be exacerbated by seasonally high sea levels.

Existing coastal sea-level products, such as NOAA's *High Tide Bulletin* and *Annual High Tide Flooding Reports* (Sweet et al., 2020), utilize characteristics of sea-level rise, tidal cycles, and vertical land motion to provide guidance about the future likelihood of coastal impacts. Yet, skillful anticipation of how sea level will evolve during the next several months or seasons remains mostly elusive. Currently there does not exist widely available monthly-to-seasonal outlooks of sea-level anomalies, as is more common for other climate variables such as sea surface temperature (Graham et al., 2011; Krishnamurti et al., 2006) and related products (e.g., coral bleaching alerts; Liu et al., 2018).

The lack of existing seasonal sea-level forecast products is not due to insufficient physical understanding of the processes that explain its variability and future evolution, nor is it due to the inability to simulate such processes in climate models. Indeed, there are a multitude of well-known sources of predictability for sea level, which include ocean dynamics and thermodynamics (e.g., Long et al., 2020; Roberts et al., 2016), all of which are incorporated into existing climate forecasting systems that utilize coupled general circulation models of the ocean and atmosphere (e.g., Kirtman et al., 2014). Many of the successes in climate forecasting are closely tied to the well-studied El Niño-Southern Oscillation (ENSO) and its seasonally dependent forecast skill (Balmaseda et al., 1995; Kumar et al., 2017; Latif et al., 1998). Just as much of the global patterns of sea surface temperature variability are explained by ENSO (Bulgin et al., 2020), so too are that of sea level (Holbrook et al., 2020).

Although sea-level forecasting has received relatively little attention compared to most other climate variables, there has been notable progress, especially in regions directly influenced by ENSO (i.e., the tropical Pacific Ocean; e.g., Widlansky et al., 2014, 2015). Miles et al. (2014) demonstrated the ability of an operational seasonal forecasting system (POAMA-2 from the Australian Bureau of Meteorology) to predict changes in satellite-based altimetry measurements of sea-level anomalies for most of the tropical Pacific. McIntosh et al. (2015) showed that the forecast skill of the same model existed for coastal locations in the tropical Indo-Pacific region, by comparing predictions to shore-based tide gauge observations. Widlansky et al. (2017) used operational seasonal forecasting models, including POAMA-2 as well as the CFSv2 model from NOAA, to demonstrate that routine prediction of sea-level variability is achievable, at least for some tropical Pacific islands where seasonal forecasts are providing coastal water level information beyond what is available in classic tide prediction calendars. The success of these studies in demonstrating the science and application of seasonal sea-level forecasting is largely because they verified model predictions against observations (i.e., from altimetry and/or tide gauges, McIntosh et al., 2015; Miles et al., 2014; Widlansky et al., 2017) that are related to coastal impacts, as opposed to relying on model simulations for verification purposes. The major limitations of these studies were the partial geographical scope (i.e., only the tropical Indo-Pacific) and the small number of models considered.

The aim of this study is to address geographic and modeling weaknesses of previous studies in assessing seasonal forecasting capabilities for sea-level variability. Specifically, we will assess the ability to predict seasonal sea-level variability for most of the world's ice-free oceans and coastlines, using a multi-model ensemble of retrospective forecasts assembled from the sea surface height (SSH) output of 10 current-generation climate prediction systems. In an investigation of the drivers of global sea-level variability, Roberts et al. (2016) showed that in one ocean eddy-permitting modeling system (NEMO) large-scale patterns of sea-level variability are predictable months-to-years in advance for many regions, at least when comparing the forecasts to the model analysis. Another study demonstrated that North Atlantic SSH anomalies are potentially predictable by using a linear inverse modeling statistical approach to assess interannual

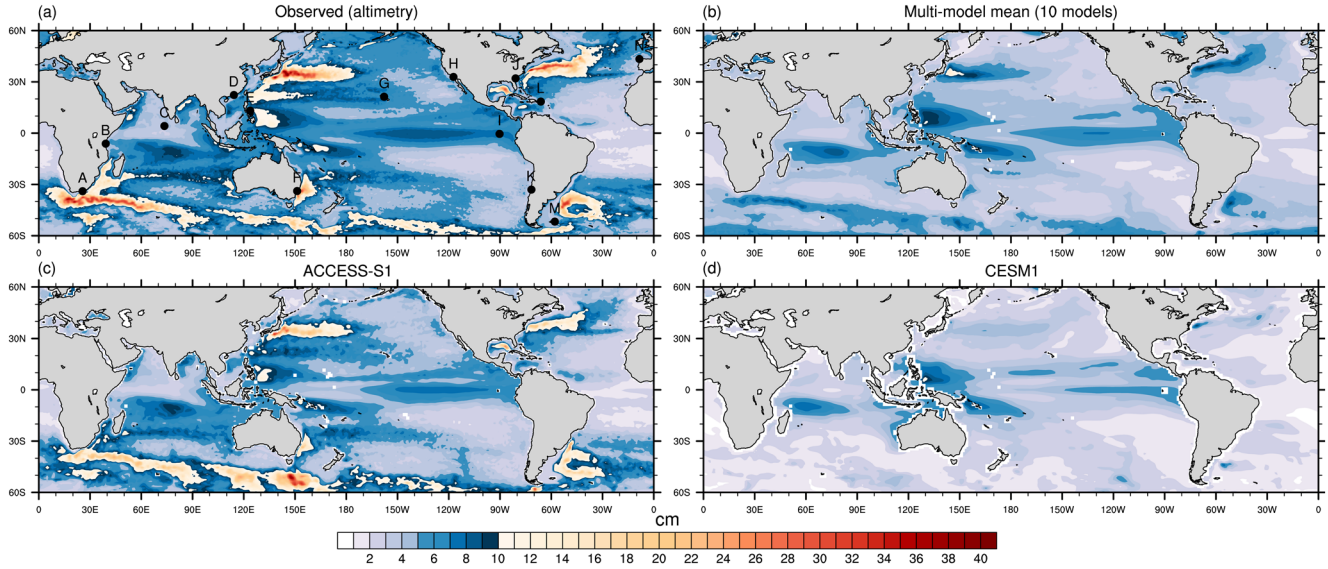


Figure 1. Observed and simulated variability of sea-level monthly anomalies (1993–2012). (a) Standard deviation of altimetry observation of sea surface height (SSH). (b) The average of standard deviations of simulated SSH for the lead-0 month forecasts from each of the 10 models (see Section 3.1). (c and d) Standard deviation of the models with either the largest (ACCESS-S1) or smallest (CESM1) variability, respectively. Only one member of each model's ensemble is used to calculate the standard deviation, which is then either averaged (b) or shown individually (c and d). Tide gauge locations used as examples are indicated by dots and labels in panel (a) (see Figure 2 for their corresponding station names).

variability in a coupled climate model (Fraser et al., 2019). Recently, Shin and Newman (2021) assessed SSH retrospective forecast skill using a similar statistical model that they applied globally, and found it to be capable of complementing the ensemble-mean forecast from five climate models that they also assessed. We will extend the work in these previous studies, primarily by comparing the forecasts to observations rather than mostly model analyses, as well as by assessing how forecast skill varies across the multi-model ensemble. Furthermore, we will determine if there is improvement in forecast skill through multi-model averaging, which may potentially cancel individual model errors (Kirtman et al., 2014). By conducting these assessments, we will test the hypothesis that skillful seasonal forecasts of monthly sea-level anomalies are possible for many coastal locations, and more fully explore the challenges and opportunities of seasonal sea-level forecasting.

Our assessment is based on comparing retrospective forecasts to observations from altimetry and tide gauges, allowing us to describe the utility of seasonal sea-level forecasting in both open-ocean and coastal locations. Section 2 describes the observations, forecasts, and skill assessment methods. Forecast skill assessments are presented in Section 3 with respect to characteristics of the modeling systems and how they are initialized, seasonal dependencies based on either when the predictions start or their target time, and how the results vary by location. Section 4 discusses the challenges of developing useful seasonal sea-level forecasts outside of the tropical Pacific, as well as opportunities for forecast system improvements. Finally, in Section 5, we conclude by outlining opportunities for meeting aspects of the societal need for future guidance on sea-level variability that, in some cases, could be readily developed using current-generation modeling systems to support new forecasting products.

2. Data and Methods

2.1. Observations

To assess the ability to forecast sea-level variability, we use monthly observations of SSH from altimetry and coastal water levels from tide gauges. We perform the analyses nearly globally (i.e., between 60°N/S) as well as for 14 locations that represent either continental or island coastlines in the tropics and midlatitudes (Figure 1a). Altimetry observations are from the SSALTO/DUACS multi-mission data set distributed by the European Copernicus Marine Environment Monitoring Service (CMEMS) on a globally uniform rectilinear

grid (0.25° resolution). We removed the Dynamic Atmospheric Correction from the CMEMS data set so that the inverse-barometer (IB) effect on sea level (Fu & Pihos, 1994; Gaspar & Ponte, 1997) is included in our observations (Figure S1 shows the monthly anomaly correlation between the IB effect and the local sea level, which is largest where the atmospheric pressure forcing on the ocean is greatest; typically in the higher latitudes). Tide gauge data, which are independent of the altimetry observations, is from the Quality Assessment of Sea Level Data archive (Caldwell et al., 2015) and likewise includes the IB effect. We only considered a sample of stations included in the UNESCO Global Sea Level Observing System's core network, and also with minimal data gaps during the time period common with the altimetry observations and sea-level forecasts (1993–2012).

Figure 1a shows the tide gauge positions with respect to the observed open-ocean patterns of sea-level variability (i.e., the sample standard deviation of monthly anomalies after removal of the annual cycle and long-term trend, as defined in Section 2.3). Sea-level variability is typically largest (i.e., standard deviations exceeding 10 cm) near the western boundary currents and associated extensions of the Pacific, Atlantic, and Indian Oceans (Hu et al., 2015), as well as in regions of strong oceanic eddy activity such as in the Southern Ocean (Faghmous et al., 2015). In the tropics, broad regions of relatively high variability are associated with ENSO (e.g., Widlansky et al., 2015) and other climate drivers influencing the Indo-Pacific region such as the Indian Ocean Dipole (Webster et al., 1999; L. Zhang et al., 2019). Standard deviations shown in Figure 1a represent the potential for either amplifying or dampening the sea-level annual cycle (Calafat et al., 2018; Vinogradov & Ponte, 2010), and thus, coastal impacts.

The comparison between altimetry and tide gauge-measured sea level is, in general, strong for the sample locations (Figure 2). There are similar amplitudes of sea-level variability observed by the tide gauges and nearest altimetry measurements (standard deviations for each are listed in Figure 2). Temporal anomaly correlation coefficients (abbreviated as ACC or r) between the two types of observations range from 0.96 at Baltra in the Galapagos Islands to 0.46 at Port Elizabeth, South Africa (respectively coded "I" and "A" on the map in Figure 1a), which are locations representative of either lower-frequency ENSO variability (Baltra) or higher-frequency eddy activity (Port Elizabeth). There is only one other location with a temporal correlation lower than 0.75 (Valparaíso, Chile; Figure 2k), which is possibly due to vertical land motion associated with earthquakes that is captured by the tide gauge but not altimetry instruments. Overall, the strong temporal coherence between altimetry and tide gauge measurements provides a consistent set of observations for assessing sea-level forecasts for both offshore and coastal regions of the world's oceans.

2.2. Retrospective Forecasts

We assessed retrospective predictions (hindcasts) of sea-level anomalies from 10 seasonal dynamical forecasting systems (Table 1). The criteria for model selection were based on the following data availability requirements: (a) forecasts for the observational period (i.e., since 1993), initialized multiple times per year (i.e., at least January, April, July, and October starts) with lead times of 6 months or longer; (b) monthly output of SSH. The forecast systems vary in design and complexity, but can be classified into three groups based on which data assimilation approaches are used to initialize the ocean component of the coupled model forecast system. Six models are from the North American Multi-Model Ensemble (NMME; Kirtman et al., 2014) which initialize the ocean using three-dimensional observations of temperature and (typically) salinity. None of the NMME models utilize altimetry observations of SSH in their ocean assimilation. Two additional models are from experiments (i.e., for nonoperational forecasting) designed to test how improving the ocean initialization affects forecast skill. These models (CFSv2-COLA and CESM1-UH; see Table 1) are initialized with temperature and salinity from the ECMWF ocean reanalysis system (ORAS4; Balmaseda et al., 2013). Altimetry observations of SSH have been used in the assimilation process for ORAS4 since 1993; therefore, such information may indirectly improve the initialization of CFSv2-COLA and CESM1-UH. The final two models (ACCESS-S1 and SEAS5; see again Table 1) assimilate altimetry-observed SSH directly into their ocean initial conditions, in addition to temperature and salinity observations. We do not assess any forecasts prior to 1993 because of the limited open-ocean observations of sea level, nor widespread measurement of subsurface temperature and salinity, which are necessary for reliable model assimilation and forecast initialization. Considering that none of the 10 models assimilate tide gauge observations, we expect that the use of altimetry data is the strongest determinant of initialization quality, potentially also

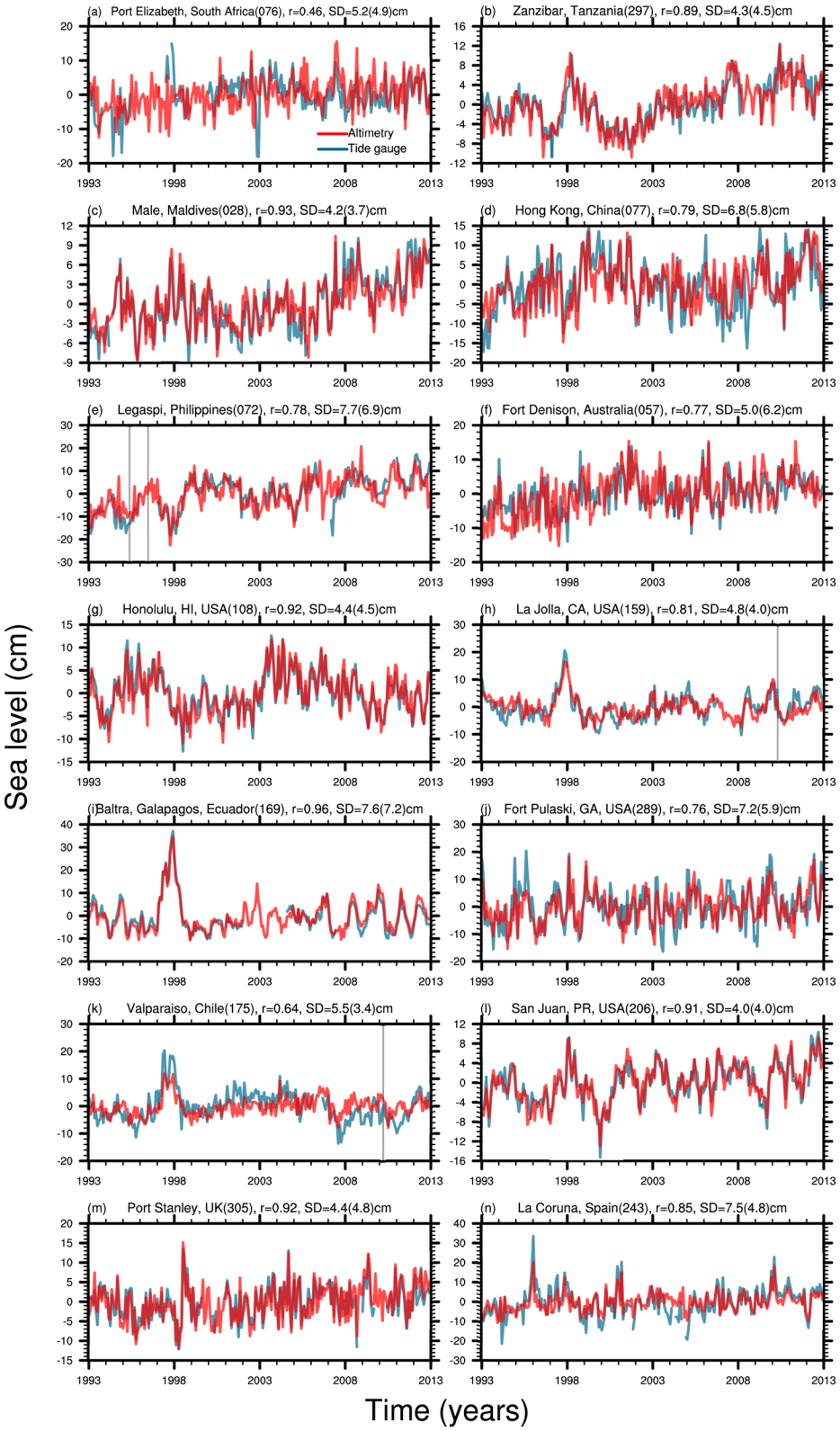

(a) Port Elizabeth, South Africa(076), r=0.46, SD=5.2(4.9)cm
(b) Zanzibar, Tanzania(297), r=0.89, SD=4.3(4.5)cm
(c) Male, Maldives(028), r=0.93, SD=4.2(3.7)cm
(d) Hong Kong, China(077), r=0.79, SD=6.8(5.8)cm
(e) Legaspi, Philippines(072), r=0.78, SD=7.7(6.9)cm
(f) Fort Denison, Australia(057), r=0.77, SD=5.0(6.2)cm
(g) Honolulu, HI, USA(108), r=0.92, SD=4.4(4.5)cm
(h) La Jolla, CA, USA(159), r=0.81, SD=4.8(4.0)cm
(i) Baltra, Galapagos, Ecuador(169), r=0.96, SD=7.6(7.2)cm
(j) Fort Pulaski, GA, USA(289), r=0.76, SD=7.2(5.9)cm
(k) Valparaiso, Chile(175), r=0.64, SD=5.5(3.4)cm
(l) San Juan, PR, USA(206), r=0.91, SD=4.0(4.0)cm
(m) Port Stanley, UK(305), r=0.92, SD=4.4(4.8)cm
(n) La Coruna, Spain(243), r=0.85, SD=7.5(4.8)cm
Altimetry
Tide gauge
Sea level (cm)
Time (years)

Table 1
Description of the Multi-Model Forecast System

Model	Organization	Ensemble size	Lead times	Resolution	Altimetry	Reference
(1) CanCM3	Canadian Meteorological Center	10	0–11	1°	No	Merryfield et al. (2013)
(2) CanCM4	Canadian Meteorological Center	10	0–11	1°	No	Merryfield et al. (2013)
(3) CCSM4-UM	University of Miami	10	0–11	1°	No	Kirtman et al. (2014)
(4) CESM1	National Center for Atmospheric Research	10	0–11	1°	No	Tribbia (2015)
(5) CFSv2	National Centers for Environmental Prediction	24 (28)	0–9	0.5°	No	Saha et al. (2014)
(6) GFDL CM2.1	Geophysical Fluid Dynamics Laboratory	10	0–11	1°	No	S. Zhang et al. (2007)
(7) CESM1-UH	University of Hawaii	5	0–11	1°	No*	Chikamoto et al. (2019)
(8) CFSv2-COLA	George Mason University	20	0–11	0.5°	No*	Huang et al. (2017)
(9) ACCESS-S1	Australian Bureau of Meteorology	11	0–6	0.25°	Yes	Hudson et al. (2017)
(10) SEAS5	European Center for Medium Range Weather Forecasts	25	0–6	0.25°	Yes	Johnson et al. (2019)

Notes. For each model, the corresponding organization, ensemble size, maximum lead (months), nominal horizontal resolution of the ocean (degrees), use of altimetry data in the forecast initialization (yes or no; asterisks note indirect information from altimetry is assimilated), and a reference are indicated. Models are ordered as six participants in the North American Multi-Model Ensemble project (1–6), two specialized experiments (7–8), and two operational forecasts that assimilate altimetry data (9–10).

affecting sea-level forecast skill. We also note that none of the models include the IB effect (see the discussion in Section 4 about opportunities for addressing this deficiency).

Horizontal resolution of the ocean models, which has been shown to affect simulated SSH variability (Penduff et al., 2010), is another differentiating characteristic of the multi-model ensemble (Table 1). Figure 1b shows the multi-model mean of the standard deviations from each of the 10 models, which contains ocean nominal resolutions of either 1° (six models), 0.5° (two models), or 0.25° (two models). For comparison of the predictions with altimetry, we regrid all of the data to a common 1° grid. Whereas, for comparison with the tide gauges, we used the nearest model data on the respective native grids. Although the global pattern of variability in the multi-model ensemble is similar to observed (Figure 1a), the magnitude of variability is substantially less in most models, especially near many continental coasts and in the higher-latitude regions that are affected by mesoscale eddies. Overall, from the 10-model ensemble, ACCESS-S1 (0.25° resolution) has the greatest variability (Figure 1c) and CESM1 (1° resolution) has the least variability (Figure 1d).

2.3. Temporal Scales and Assessment Metrics

We assess monthly and seasonal (3-month average) SSH anomalies from the forecast models for all lead times (0–11 months), focusing on the 6-month outlook because that is the longest lead available for all models (Table 1). In calculating the anomalies, for each model, we first remove the respective initial time and lead-time dependent climatology (1993–2012), which is necessary to bias correct for near-term model drift (Smith et al., 2013; Vannitsem et al., 2018). Lastly, we remove any long-term trends from the observations

Figure 2. Observed sea-level monthly anomalies during 1993–2012 for 14 locations sampling the global continental and island coasts. Tide gauge (blue) and altimetry (red) data are ordered consistent with the location lettering in Figure 1 (station names and Global Sea Level Observing System identification numbers are indicated above each panel). The altimetry data are interpolated to the tide gauge locations by using the nearest grid point within the original 0.25 grid. Anomaly correlation coefficient values (r) between the tide gauge and altimetry are listed along with the standard deviation (SD) of the tide gauge and altimetry (in parentheses). Times of earthquakes occurring within 200 km of a location and exceeding 7.5Mw are indicated by vertical lines, which may indicate potential vertical land motion.

and forecasts (i.e., the linear trend over the retrospective period; Figure 3), which we calculate for the latter based on the lead-0 month anomalies of each model.

The importance of removing long-term trends from the forecasts, which we perform by grid cell, is due to the large differences in both global and regional trends across models, and also, compared to observations (Figure 3 and Table 2). In general, all the ocean models use the Bousinesq approximation, which implies that the total ocean volume can not change in response to global steric changes. The models will capture variations in the steric component, however, and will also respond to changing mass from freshwater input. Only two forecast systems are designed to incorporate global sea-level rise (ACCESS-S1 and SEAS5), because their assimilation systems include special treatment of altimetry (Balmaseda et al., 2013). Much smaller global mean sea-level trends in the CFSv2 and GFDL models are presumably caused by their assimilation procedures, although this is not yet fully understood, whereas the other six models have no trend. Regionally, all of the models differ to some extent from the altimetry-observed long-term trend (Figure 3a), although biases are typically largest in the models that do not assimilate altimetry (e.g., there are pronounced negative trends in the North Atlantic for several of the NMME models including CFSv2; Table 2). In models that do not include global sea-level rise, removal of the long-term trend improves overall forecast skill by reducing regional biases, which in many places are comparable to the magnitude of observed variability (Widlansky et al., 2017). Furthermore, trend removal also helps to achieve an equitable multi-model assessment of seasonal forecasting capabilities (i.e., by treating models equally regardless of how global sea-level rise is resolved). Potential consequences of not removing long-term trends are discussed in Section 4.

We assessed the forecast skill of individual models and the multi-model ensemble using metrics of the ACC, as well as the root-mean-square error (RMSE), which were both calculated between observations and the respective predictions. ACC measures how well the model can predict the phase of observed variability, whereas RMSE measures the magnitude of amplitude errors (Wilks, 2011). The RMSE values are compared to the observed standard deviations (Figures 1a and 2). Statistical significance of the ACC values is based on a one-tailed t-test against a null hypothesis of zero correlation. The degrees of freedom of the significance test are determined by the observed autocorrelation decay timescale at each location (i.e., $N^* = N\delta t/2\tau$, where N is the number of observations, δt is the sample interval, and τ is the e-folding time). Note that because of the short sample of retrospective forecasts (i.e., 20 years, and typically $N^* = 20$ for testing the significance of forecast skills from individual start times), we do not test significance of the correlation differences between models.

We also compared the ACC values from the 10 dynamical models to those from a damped-persistence statistical model of SSH (Figure 4 and Figure S2, respectively, for monthly and seasonal anomalies). The damped-persistence model was constructed by persisting observed monthly anomalies through all forecast leads, but damped at a rate equal to the autocorrelation of the anomalies. We note that the ACC of a damped-persistence model is the same as simple persistence, except for locations with a negative autocorrelation at a particular lead time, which gives a minor skill advantage to including the damping timescale. For the damped-persistence model, we use the altimetry or tide gauge observed SSH anomaly from the month prior to the forecast initialization month (e.g., December of the prior year is used for January starts of both the monthly and seasonal forecasts, with the lead-1 month corresponding to February and the lead-1 season to January–March). Note that we defined the leads of the seasonal forecasts in such a way that the middle months match the monthly forecasts. Since the ACC between altimetry observations and the damped-persistence forecast of monthly and seasonal sea-level anomalies is statistically significant in parts of the world's oceans out to at least six months (e.g., in part of the tropical central Pacific; Figure 4 and Figure S2), determining whether the dynamical models are more skillful than this simple statistical model is an additional assessment of the multi-model forecasting utility.

3. Results

3.1. Initialization Quality and Forecast Skill

Using a subset of three models (CFSv2, CFSv2-COLA, and SEAS5) as an example of the widely ranging differences in initialization quality in the multi-model ensemble, we find a disparity in how well the models initially capture observed monthly sea-level anomalies. Figures 5a–5c shows, for these three models, the

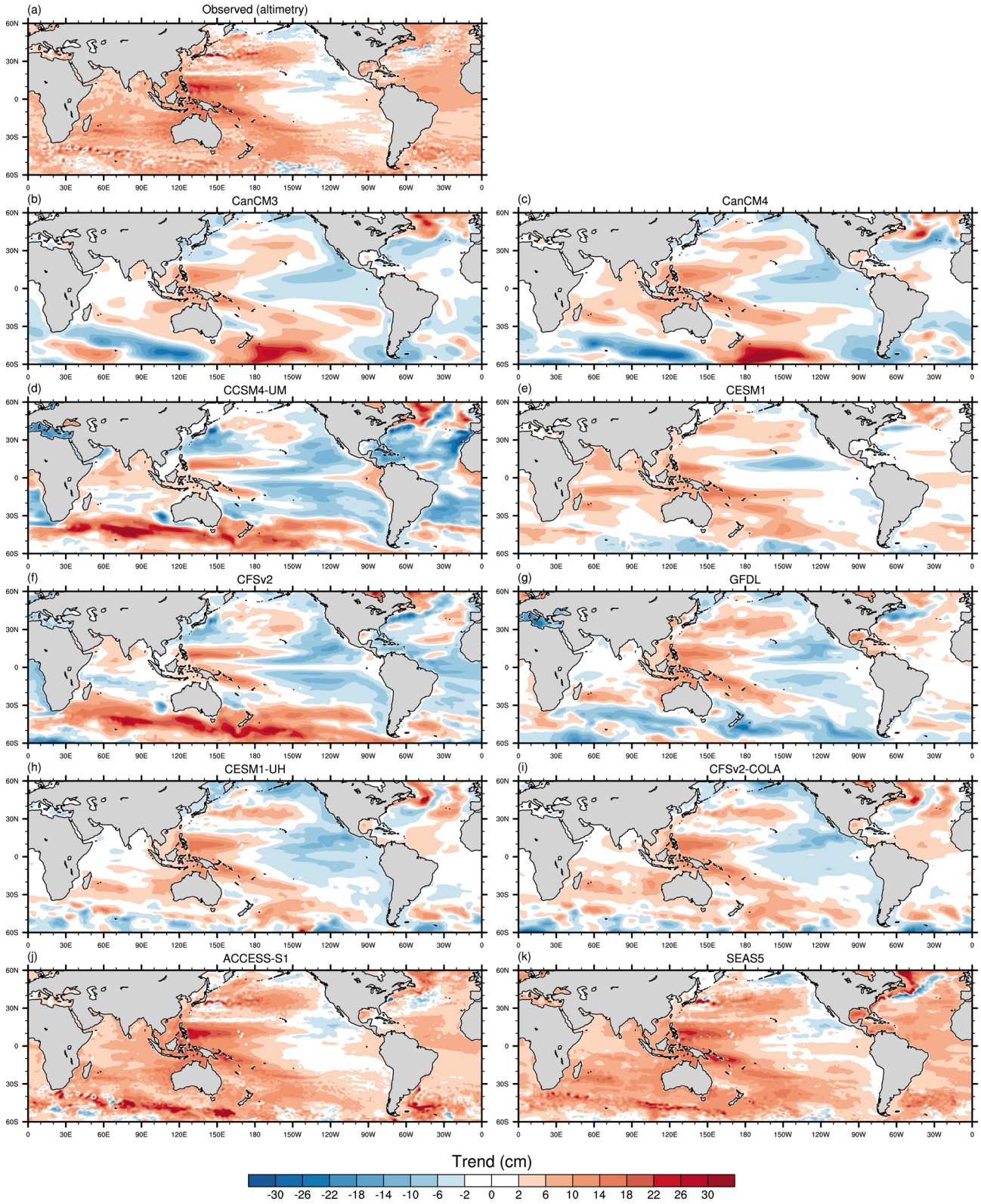


Figure 3. Long-term trends of observed and simulated sea-level monthly anomalies. Linear trends (cm) are calculated for the 20-year epoch (1993–2012) for the altimetry (a) and each of the 10 models at the lead-0 month (b–k). Table 2 lists the global (60°N/S) and North Atlantic (0°–60°N) regional averages of the trends.

Table 2
Linear Trends (cm; 1993–2012) Averaged Over the Global and North Atlantic Domains (60°N/S and 0°–60°N, Respectively) Shown in Figure 3, and the Global Averages of Standard Deviation (Global SD)

	Global trend	North Atlantic trend	Global SD
Altimetry	2.89	2.86	6.4
CanCM3	0.00	−0.49	3.0
CanCM4	0.00	−0.20	3.2
CCSM4-UM	0.00	−2.10	4.9
CESM1	0.00	−0.10	2.8
CFSv2	0.32	−1.10	4.6
GFDL	−0.07	0.63	3.5
CESM1-UH	0.00	0.41	3.5
CFSv2-COLA	0.00	0.91	4.0
ACCESS-S1	2.85	2.19	5.5
SEAS5	2.05	2.29	4.7

Notes. The observed and assimilated trends standard deviation are listed, respectively from altimetry and the 0-month lead forecast of each model. A single member of the model output is used to calculate the standard deviation for each model.

ACC between altimetry observations and the lead-0 month forecast; the latter of which approximates the model initialization or analysis of the observed sea-level anomaly (Figure S3). The initialization quality appears to be most influenced by which ocean data is assimilated (Table 1; see also Figure S4 for ACC assessments of the other seven models), in particular whether or not altimetry is utilized.

The CFSv2 model does not assimilate altimetry, nor do the five other models participating in the NMME, as we mentioned in Section 2.2. At the lead-0 month, CFSv2 exhibits the lowest correlations with observations nearly everywhere outside of the tropical Pacific (Figure 5a), at least compared to the two other models in the subset (Figures 5b and 5c). Throughout most of the Atlantic Ocean, as well as in the higher-latitude Indian and Pacific Oceans, the correlation between observations and CFSv2 are not statistically significant, despite bias corrections such as removing the long-term trend (Figure 3f), which is especially erroneous in the North Atlantic (Table 2). The CFSv2-COLA experiment demonstrates that by improving the ocean assimilation quality (see Section 2.2 and Table 1), the lead-0 month correlations with observations are improved in the tropics and subtropics (including most of the Atlantic Ocean), although correlations remain low in the midlatitudes (Figure 5b). In contrast especially to CFSv2, the SEAS5 model has statistically significant correlations at lead-0 month that extend further poleward in all ocean basins (Figure 5c). SEAS5 also has the highest correlations in the tropics at lead-0 month, especially in comparison to CFSv2. We note that, for all of the models, the ACC is still lower in the tropical Atlantic where observed variability is small (Figure 1a), compared to similar latitudes of the Pacific and Indian Oceans (Figures 5a–5c and Figure S4). In general, we suspect that the assimilation of altimetry into the ocean initialization by SEAS5 (together with ACCESS-S1, which uses a similar assimilation method; see that model's ACC in Figure S4) is contributing to the higher correlations in most regions. However, the higher horizontal resolutions in both of these models may also possibly be improving how well the finer-scale sea-level variability in the midlatitudes is resolved (cf. Figure 1c).

Correlations between the lead-6 month forecasts and observations (Figures 5d–5f; Figure S4) are generally similar in pattern to those at the lead-0 month. As expected though, skill degrades with increasing forecast lead time for all models (i.e., correlations are lower for the lead-6 month forecast compared to at the lead-0 month). At 6 months lead, all of the models predict the observed variability better in the tropics compared to the midlatitudes. The correlations also tend to be much higher in the tropical Pacific compared to the tropical Atlantic, similar to the lead-0 month pattern. However, the models with initial conditions closer to that observed for the Atlantic Ocean (e.g., CFSv2-COLA and SEAS5; Figures 5b and 5c) also have the highest correlations there at lead-6 months (Figures 5e and 5f), especially in the tropics and subtropics. In the midlatitudes of each ocean basin, the lead-6 month forecast skill is similarly poor across all models (i.e., correlations with observations are rarely statistically significant; Figures 5d–5f and Figure S4), as is also mostly the case for those regions at the lead-0 month (Figures 5a–5c). We note that the lower skill in the midlatitudes is partly because we are validating the lead-0 and -6 month forecasts against altimetry (and tide gauge) observations that include the IB effect, even though none of the models simulate atmospheric pressure forcing on the ocean surface (for indication of where this is most likely to be a deficiency, Figure S1 shows the global ACC pattern of the IB effect with monthly sea-level observations).

The relationship between initialization quality and forecast skill (e.g., at lead-6 months) applies to the global oceans (Figure 6a) and to the coastal tide gauge locations (Figure 6b). Considering the 10 models individually (circles in Figure 6a), correlations at lead-0 and -6 months at a particular grid point are somewhat linearly related (i.e., the regression slope is close to 1). The linear relationship between lead-0 and -6 month correlations holds when considering the multi-model mean forecast for all global locations (shading in Figure 6a), as well as at the tide gauges (Figure 6b). On a global-average basis, the multi-model mean forecast

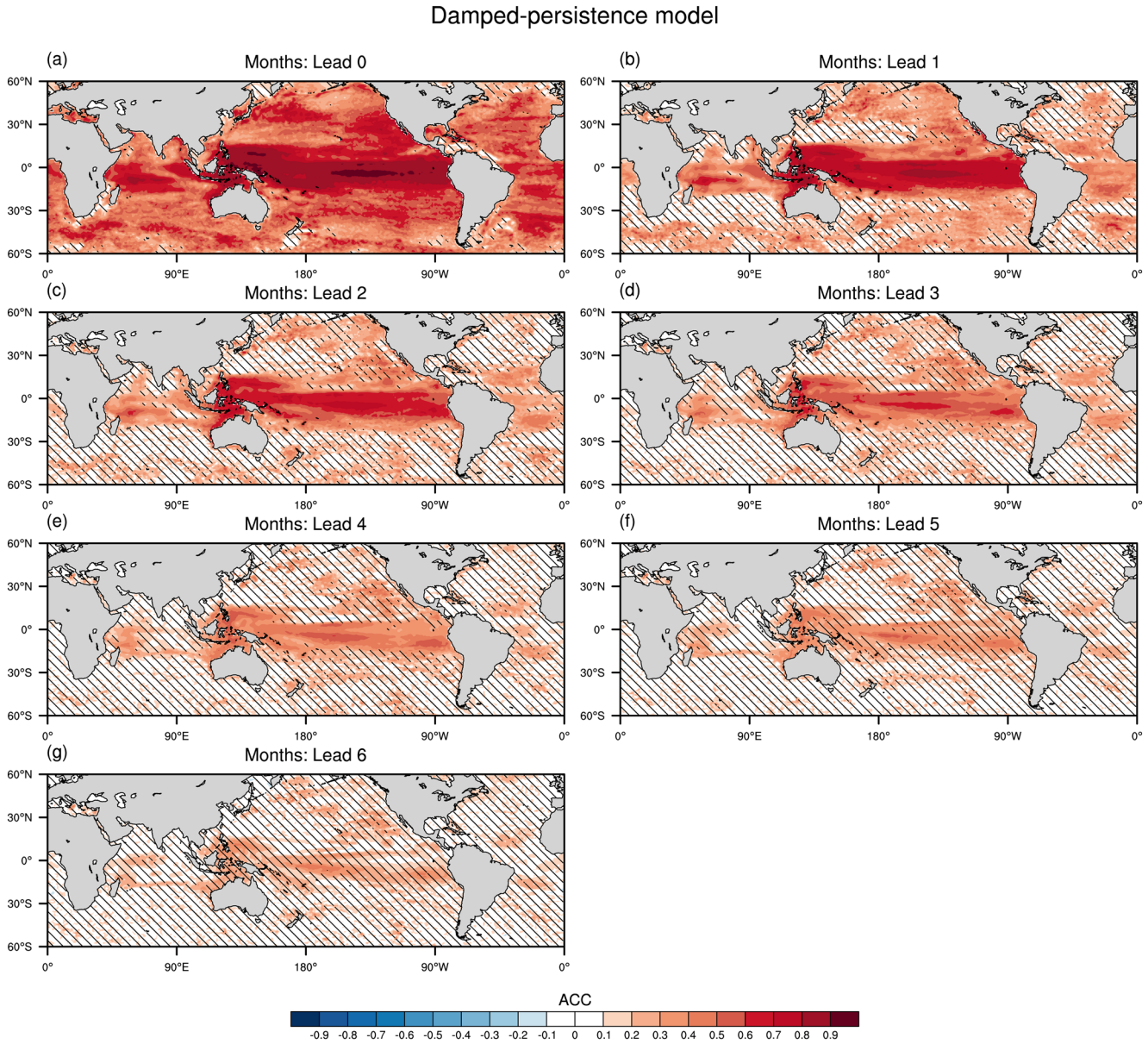


Figure 4. Retrospective forecast skill for the damped-persistence model compared to altimetry. Anomaly correlation coefficient values are shown for lead-0 to -6 months (a–g, respectively). All four start times (January, April, July, and October) are considered. Hatching indicates correlations that are not statistically significant at the 5% level using a one-tailed *t*-test.

at lead-0 and -6 months typically performs better than the individual models (triangle compared to circles in Figure 6a). If we assess the performance at each location over the globe instead of the global average skill, the multi-model mean forecast also performs well. In fact, at the 0-month lead, ACCESS-S1 is the only model that performs better than the multi-model mean in a majority of the global area that we analyzed (orange dots above the one-to-one line in Figure 7), which is presumably related to its assimilation of altimetry observations (Table 1). The next three best performing models compared to the multi-model mean also assimilate altimetry, either directly (SEAS5) or indirectly (CFSv2-COLA and CESM1-UH). At the 6-month lead, none of the 10 models perform better than the multi-model mean forecast in a majority of the globe (blue dots in Figure 7), although according to this metric SEAS5 is the top-ranked model and ACCESS-S1, CFSv2-COLA, and CESM1-UH are all in the top half (they are joined by CESM1, which is also in the upper half of the ensemble at lead 0 despite not assimilating any altimetry information).

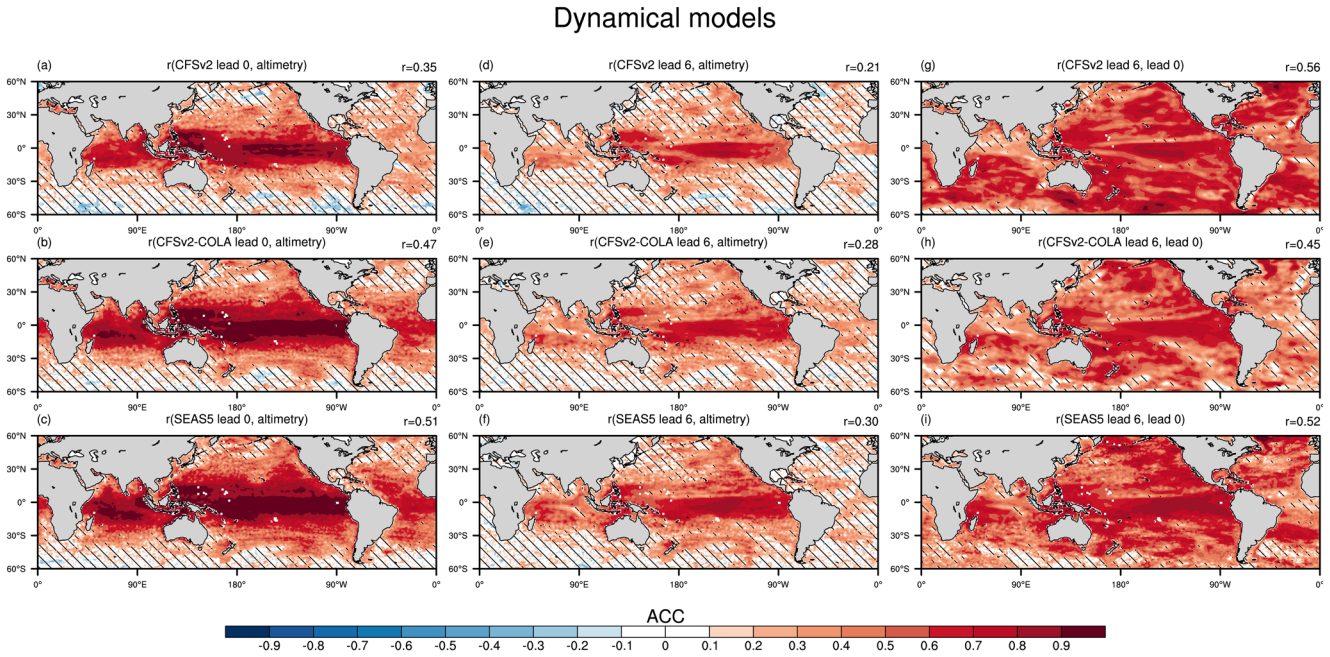


Figure 5. Retrospective forecast skills at lead-0 and -6 months for three dynamical model forecasts of monthly sea-level anomalies. These models are chosen from the 10-model ensemble to represent ocean initializations utilizing either no altimetry observations (CFSv2), temperature and salinity that are influenced by altimetry (CFSv2-COLA), or assimilation with altimetry (SEAS5). Anomaly correlation coefficients (ACCs) are between the model forecasts and either the altimetry (a–c for lead-0 month, and d–f for lead-6 months) or a single member of respective model lead-0 month forecast (g–i for lead-6 months). All four start times (January, April, July, and October) are considered. Hatching indicates correlations that are not statistically significant at the 5% level using a one-tailed *t*-test. The domain average of ACC is shown for each panel (*r*).

We have so far validated the forecasts using observed sea level, however it is also informative to compare the forecasts to the sea-level anomalies that are predicted by each of the models soon after the initialization process (i.e., assessing correlations between the lead-6 month forecast and the lead-0 month anomalies

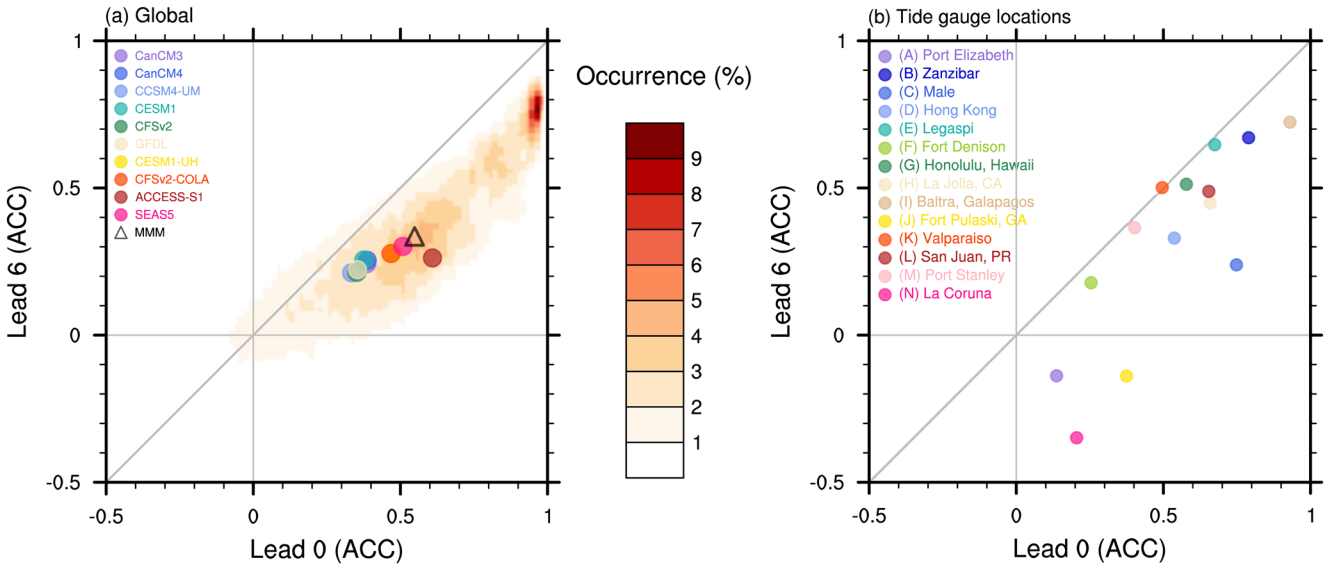


Figure 6. Global (a) and local (b) retrospective forecast skill at lead-6 months as a function of the skill at lead-0 month. Skill is calculated at each location as the anomaly correlation coefficient (ACC) between the forecast and either altimetry (a) or tide gauge (b) observations. (a) Kernel density estimate (% of correlation values) at each grid point from the multi-model mean forecast (shading; probabilities are multiplied by 10^2); each colored circle represents the domain average of ACC skills for one model, and the black triangle represents the multi-model mean (MMM). (b) Scatter plot of correlation values of multi-model mean forecast for the 14 sample locations shown in Figure 1 (letter codes are in parentheses). Diagonal lines represent a regression slope of 1.

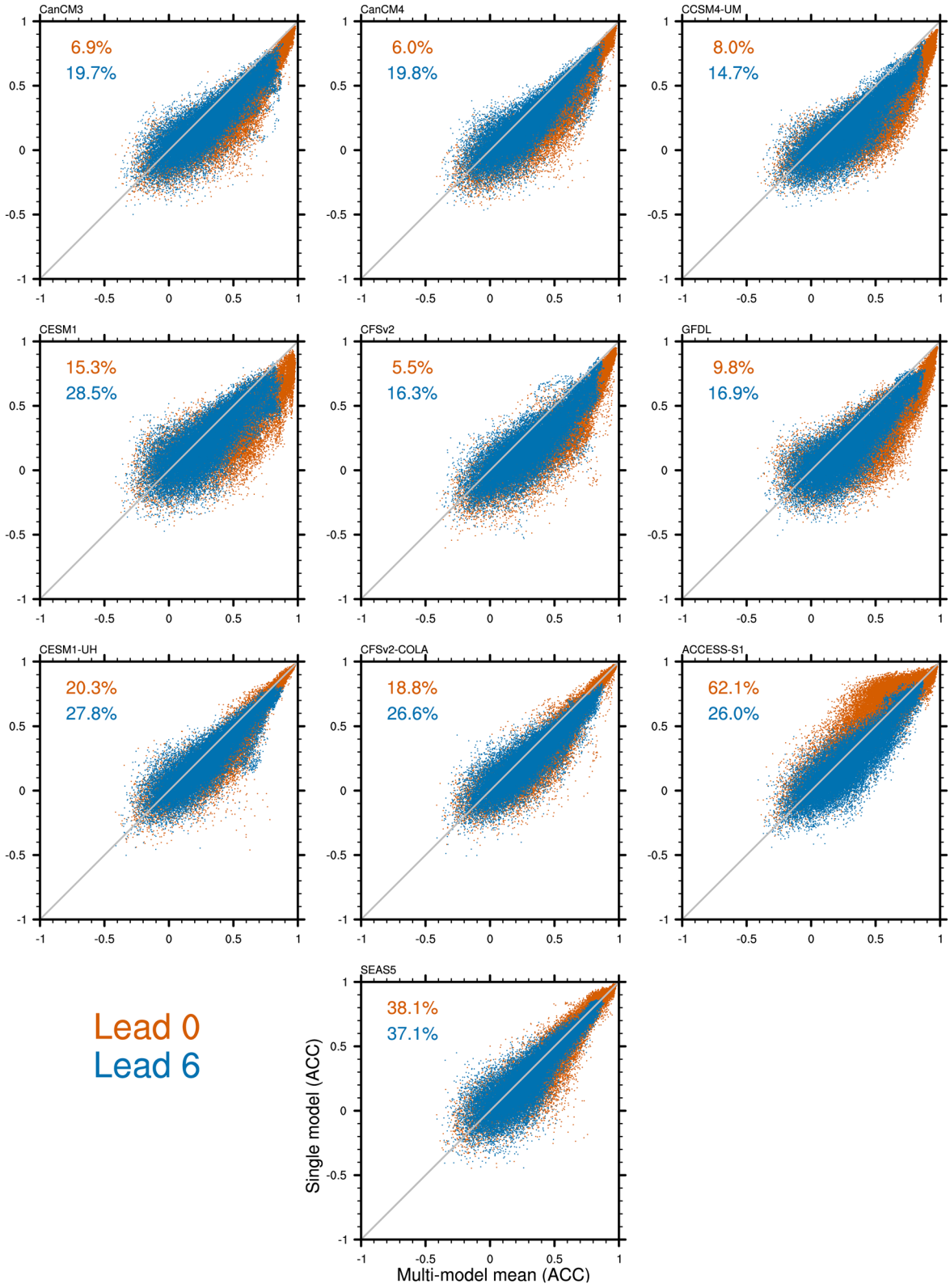


Figure 7. Retrospective forecast skills at every grid point for each of the 10 models (*y* axes) and the multi-model mean forecasts (*x* axes are consistent in all panels). The anomaly correlation coefficient skill is calculated for all grid points in the domain (60°N/S), for lead-0 month (orange) and lead-6 months (blue). In each panel, the percentage of grid points where the particular model has a higher skill than the multi-model mean is indicated.

corresponding to that time). There is a dual purpose of such a comparison: first, it informs as to what forecast skill could be realized if a model fully resolved the observed sea-level variability and, second, skill assessments of seasonal sea-level forecasts often use the model analysis (i.e., initial conditions) as observations (e.g., Roberts et al., 2016). Note that we use the lead-0 month forecast as an approximation of a model's analysis (cf. the high ACC between lead-0 month of either CFSv2 or SEAS5 and, respectively, the Climate Forecast System Reanalysis or Ocean Reanalysis System 5; Figure S3). The downside of using the model analysis (or lead-0 month forecast) to assess forecast skill is obvious. An inflation in correlations is likely, compared to what would be expected in a real-time and practical situation of comparing forecasts with observations. As expected, the correlations between lead-0 and -6 month forecasts are typically much higher compared to the assessment using observations (Figures 5g–5i vs. Figures 5d–5f). However, the difference in lead-6 month skill based on either a verification using the lead-0 month or altimetry is less for models that initialize more closely to observations (e.g., comparing CFSv2 and CFSv2-COLA; see also Figure S4, which shows the smallest difference between the choice of verification is for ACCESS-S1, presumably because that model's lead-0 month has the highest correlation with altimetry). When compared to the respective lead-0 month anomalies, the lead-6 month forecasts are skillful nearly everywhere (i.e., statistically significant ACC). In the Atlantic Ocean, comparing the forecasts against the lead-0 months (i.e., approximately the analyses) diminishes the skill disparity across the three example models (CFSv2, CFSv2-COLA, and SEAS5) and among most of the remainder of the multi-model ensemble (Figure S4). We discuss in Section 4 implications of the higher correlations between forecasts and model analyses, but otherwise the actual observations (either from altimetry or tide gauge measurements) will be used for all further assessments of forecast skill, with our focus being to measure the performance of the multi-model mean.

3.2. Seasonal Dependence of the Forecast Skill

Forecast skill at a particular lead time is dependent on both the model and forecast location, as well as the initialization and target months (Huang et al., 2017; Xue et al., 2013). Analyzing the retrospective sea-level forecasts based on the initialization month (i.e., January, April, July, or October) reveals seasonally dependent differences in skill in many parts of the world's oceans (Figure 8). In the near-equatorial Indo-Pacific region, the January initializations of the 6-month forecasts (July targets; Figure 8a) tend to be less skillful compared to the July starts (January targets; Figure 8c). Comparing 3-month average forecasts (Figure S5) started from similar initialization months (i.e., target seasons of June–August vs. December–February) shows the same tendency for July forecast starts to be more skillful, at least for this lead time and in regions most directly affected by ENSO (i.e., a manifestation of comparing forecasts initialized before or after the so-called ENSO spring predictability barrier; e.g., Torrence & Webster, 1998). The seasonal (3-month average) forecasts are also somewhat more skillful overall, compared to the monthly forecasts.

Improvement of the multi-model mean forecasts, compared to the damped-persistence forecasts initialized from corresponding months (Figure 8, right column), also reveals a pattern that changes throughout the year. However, the pattern of the multi-model forecast improvement is somewhat counter intuitive because the persistence of sea-level anomalies also varies seasonally. In locations strongly influenced by ENSO (e.g., the Galapagos Islands; cf. Figure 2i), sea-level anomalies are less persistent around the typical peak of El Niño or La Niña events (i.e., December or January, such as during 1997/1998 or 1998/1999, respectively), as compared to those during the canonical development phase of ENSO that occurs in boreal summer. Most likely due to this phase locking of ENSO to the annual cycle (e.g., Chen & Jin, 2020), dynamical model forecasts initialized in January (July) tend to perform better (worse) than persistence in the equatorial eastern Pacific (Figures 8b and 8f). If the multi-model forecast is assessed simply compared to the observations (Figures 8a and 8e), the opposite seasonality pattern of skill is apparent.

Outside of the equatorial Pacific, there are some locations where seasonal dependence of forecast skill also exists, and other places where skill (or lack thereof) is more consistent throughout the year. Sea-level anomalies along the US West Coast in January (Figure 8e) are more accurately predicted than in July (Figure 8a) at 6 months lead by the multi-model mean forecast, which is likely another example of how the models' ability to resolve ENSO dynamics affects the forecast performance. In contrast, near the US East Coast and elsewhere in the Atlantic Ocean, there is no apparent seasonality of the forecast skill, at least at these broad scales (Figure 8 and Figure S5).

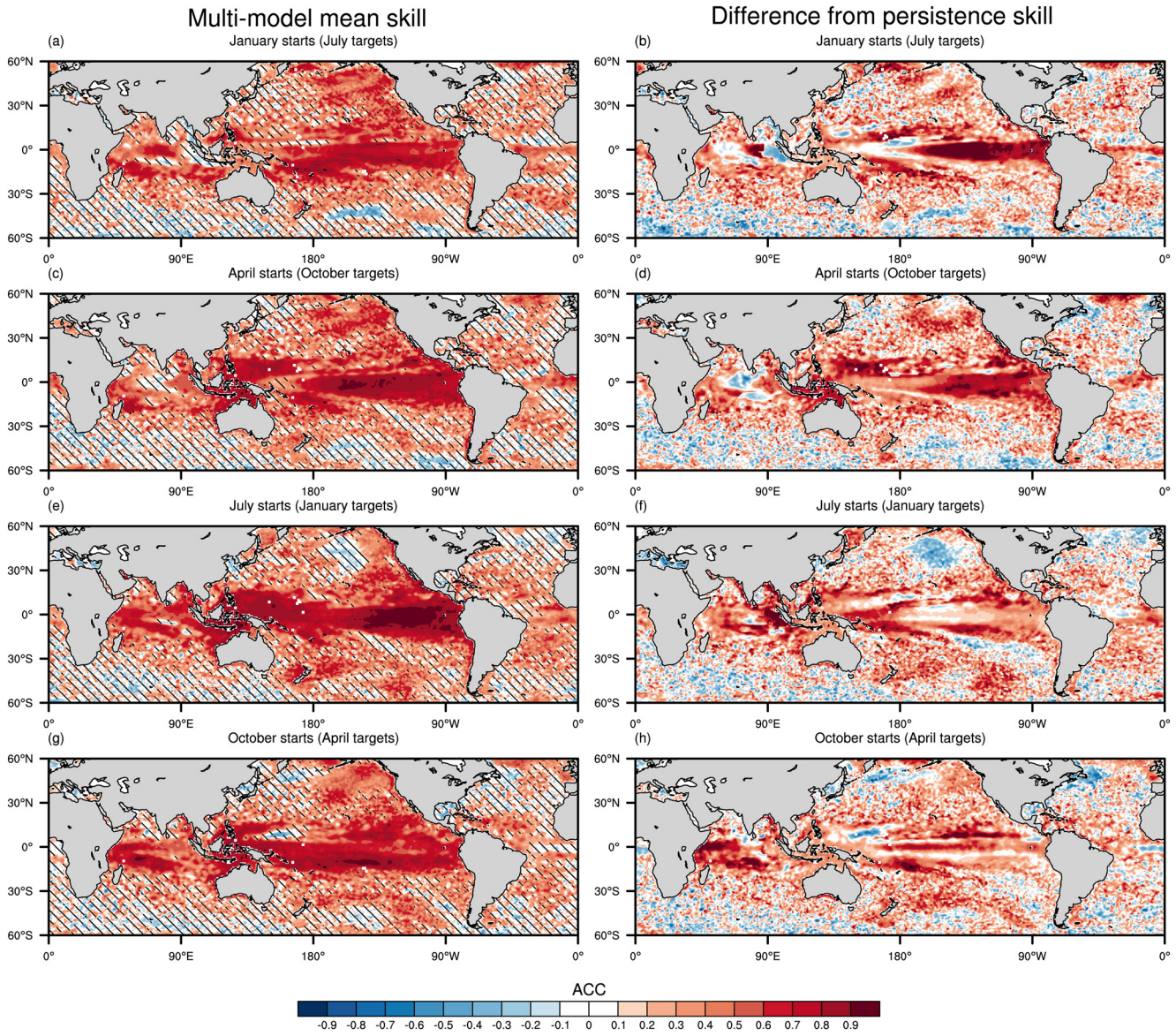
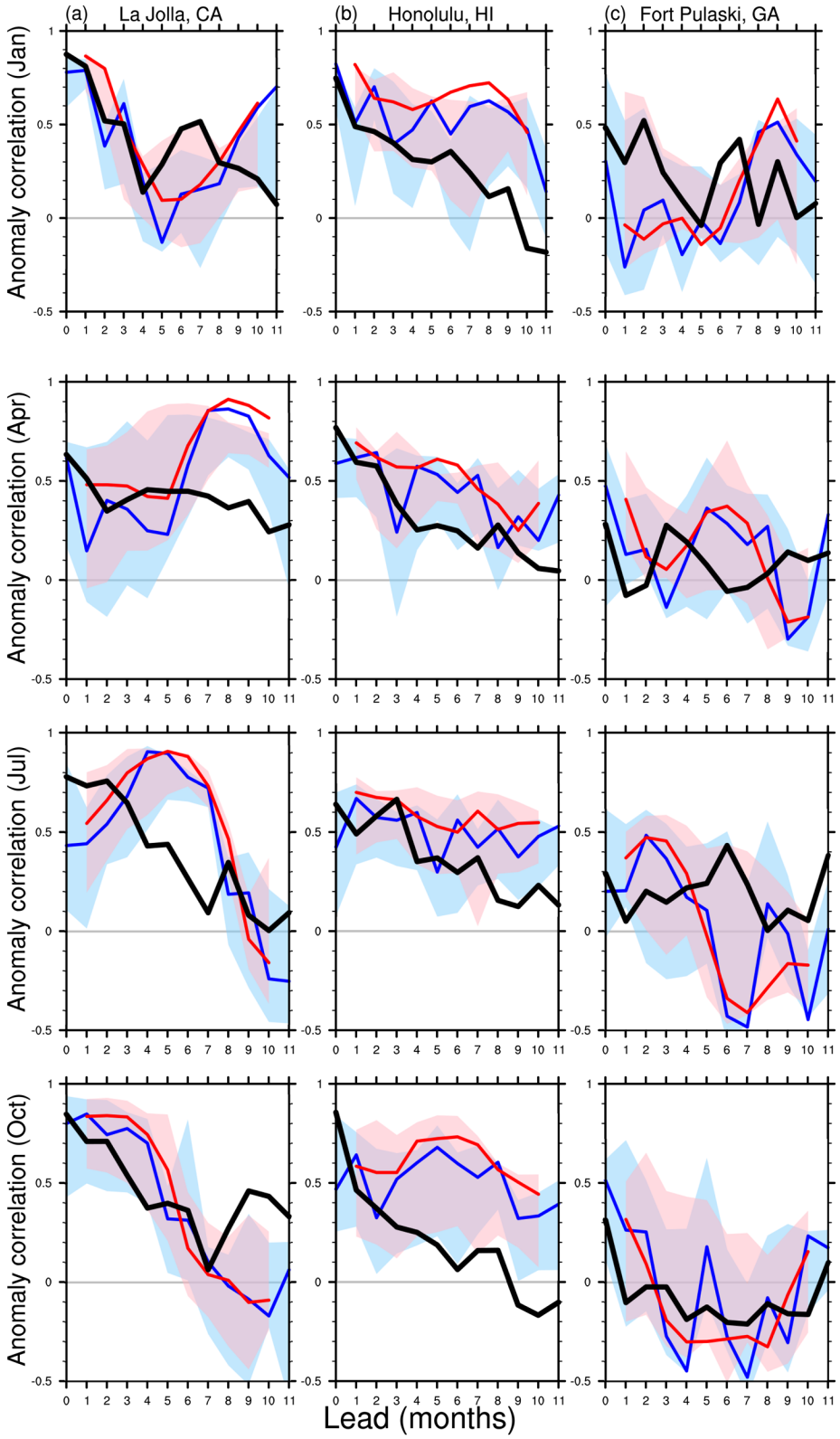


Figure 8. Retrospective forecast skills (anomaly correlation coefficient [ACC]) at lead-6 months for the multi-model mean forecasts of monthly sea-level anomalies. Forecasts are initialized during January (a), April (c), July (e), or October (g) and compared to altimetry observations during the following July, October, January (year +1), or April (year +1), respectively. Hatching indicates correlations that are not statistically significant at the 5% level using a one-tailed *t*-test. (b, d, f, and h) Differences between the ACC (forecasts compared to altimetry) using the multi-model mean or the damped-persistence model.

We also assessed the seasonality of forecast skill for specific coastal locations, using tide gauge measurements as observations (Figure 9 and Figure S6). Unlike the global assessment, for the coastal locations we compared the multi-model mean forecast to the inter-model spread (i.e., in Figure 9, we group the forecasts by initialization month and consider all available lead times for each model as noted in Table 1). Three locations (La Jolla in California, Honolulu in Hawaii, and Fort Pulaski in Georgia) provide examples of different characteristics in forecast skill. Forecasts of sea-level anomalies at La Jolla are clearly the most seasonally dependent (Figure 9a), similar to near the rest of the US West Coast (Figure 8). Regardless of the initialization month, forecasts for the upcoming boreal winter months (December–February target; e.g., lead-6 months from the July start or lead-3 months from the October start) at La Jolla have the highest correlations with the tide gauge observations (typically, ACC > 0.7). For Honolulu (Figure 9b) in the north-central tropical Pacific, the seasonal dependence is much smaller, with the lead-6 month correlations from the four different start months all between 0.4 and 0.6. The forecasts for Fort Pulaski on the US East Coast


(a) La Jolla, CA
(b) Honolulu, HI
(c) Fort Pulaski, GA
Anomaly correlation (Jan)
Anomaly correlation (Apr)
Anomaly correlation (Jul)
Anomaly correlation (Oct)
Lead (months)

(Figure 9c), compared to either La Jolla or Honolulu, have even lower skill for nearly all start and target months; however, a subtle peak of correlations occurs there during boreal fall, regardless of when the forecast started (e.g., for lead-9 months after the January start, ACC ≈ 0.5). In general, the multi-model mean correlations at Fort Pulaski are rarely higher than those for the damped-persistence model, which is an example of forecasting challenges faced at some other coastal locations as well (see Figure S6 and Section 3.3).

The analyses of forecasts compared to tide gauge observations (Figure 9 and Figure S6) further elucidate two characteristics of the multi-model mean forecasts. First, for all of the sample locations, the 3-month average forecast compared to the monthly forecast, is typically similar, or sometimes somewhat improved. Second, the correlations of the multi-model mean forecasts are usually within the inter-model range at most leads. The first characteristic was noted previously for lead-6 month forecasts, such as when considering the global correlation maps (Figure 8 and Figure S5; monthly vs. 3-month average forecasts). In contrast, the second characteristic (i.e., skill of the multi-model mean forecast does not usually beat all of the individual models at the sample locations; Figure 9) seems to conflict with the global assessment showing that for most places the multi-model mean forecast performs best (Figure 7). Implications of these characteristics concerning the forecast timescale (i.e., monthly vs. seasonal targets) and inter-model spread (i.e., individual forecasts vs. multi-model averaging), with regards to forecasting improvements, will be discussed in Section 4.

3.3. Analysis of Coastal Forecast Skill

The global assessment of retrospective sea-level forecasts (e.g., Figure 8) revealed vast regions that are skillful out to at least 6 months, when compared to altimetry observations. Statistically significant correlations, as well as forecasts that are better than if using simply a damped-persistence model, are especially widespread in the tropical Indo-Pacific but also in parts of the tropical Atlantic basin (e.g., around the eastern Caribbean Sea) and in a few midlatitude locations of each ocean. In contrast, assessing the forecasts based on comparison to tide gauge observations (Figure 9 and Figure S6) identified coastal locations with widely varying forecast performances (e.g., high skill at La Jolla vs. low skill at Fort Pulaski, which are characteristic of the US West and East Coasts, respectively). Here, we will explore the performance of coastal forecasts for both of these places, as well as Honolulu to provide an island example (results for the other coastal locations are in the Supplementary Information).

Point-wise correlations between the tide gauges and regional altimetry observations reveal that coherence typically extends far away from the coastal location (Figures 10a–10c and Figure S7), though the correlation spatial patterns differ between regions. When the Honolulu tide gauge records above-normal sea levels, altimetry observations are also typically high not only throughout the Hawaiian Islands but also in a broad region extending over 1,000 km to the east (Figure 10a), which is a pattern that has been shown to be related to oceanic Rossby wave propagation and air-sea thermodynamic forcing (Long et al., 2020). Altimetry observations are also highly correlated with the La Jolla and Fort Pulaski tide gauges ($r = 0.81$ and $r = 0.76$, respectively, compared to $r = 0.92$ at Honolulu; Figure 2), with patterns extending far from the tide gauges along the coast (i.e., encompassing the entire US West and Southeast Coasts, respectively; Figures 10b and 10c). The former pattern has been linked to coastally trapped Kelvin wave propagation from the equatorial Pacific (Allen, 1975). Whereas the latter pattern is possibly explained by how westward propagating Rossby waves interact with the continental slope, and thereby, affect coastal sea levels from Cape Hatteras southward into the Gulf of Mexico (Calafat et al., 2018).

For the multi-model lead-0 month forecasts, correlations between the local and regional sea-level anomalies around Honolulu, La Jolla, and Fort Pulaski (Figures 10d–10f) are similar spatially to the correlation patterns between the tide gauges and altimetry observations (Figures 10a–10c). However, compared to the observed patterns, the correlations for lead-0 month (i.e., the patterns internal to the forecast models) are

Figure 9. Retrospective forecast skill as compared to tide gauge observations (anomaly correlation coefficient on the *y*-axes) for three sample locations: Honolulu (a), La Jolla (b), and Fort Pulaski (c). Forecasts are initialized during January (row 1), April (row 2), July (row 3), or October (row 4) and extend from lead-0 to 11 months (*x*-axes). Skill of the multi-model mean forecasts of monthly (blue) and seasonal (red) sea-level anomalies are indicated by lines. The respective color shadings indicate inter-model spread of skill at forecasting the monthly and seasonal anomalies. For comparison, skill of the damped-persistence forecast of monthly anomalies are indicated by black lines. Note that all leads are unavailable for three of the 10 models (see Table 1).

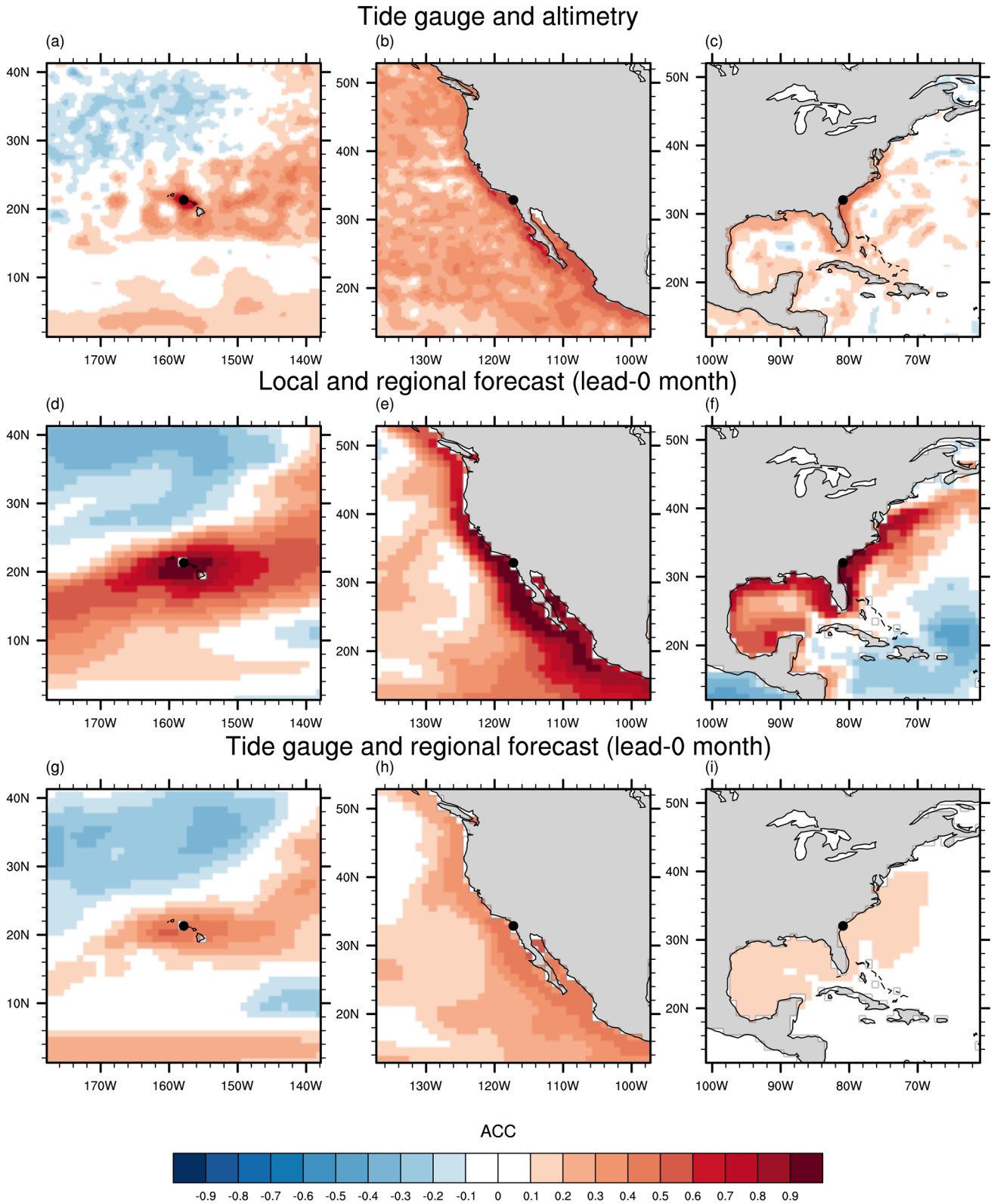


Figure 10. Spatial correlation patterns between sea-level anomalies at three sample locations (Honolulu, La Jolla, and Fort Pulaski) and grid points within 40 latitude-longitude regions centered on each location. Three correlation maps are shown for each location: (a–c) tide gauge and altimetry observations; (d–f) local forecast and regional forecast; (g–i) tide gauge observation and regional forecast. The forecasts are for lead-0 month using the multi-model mean.

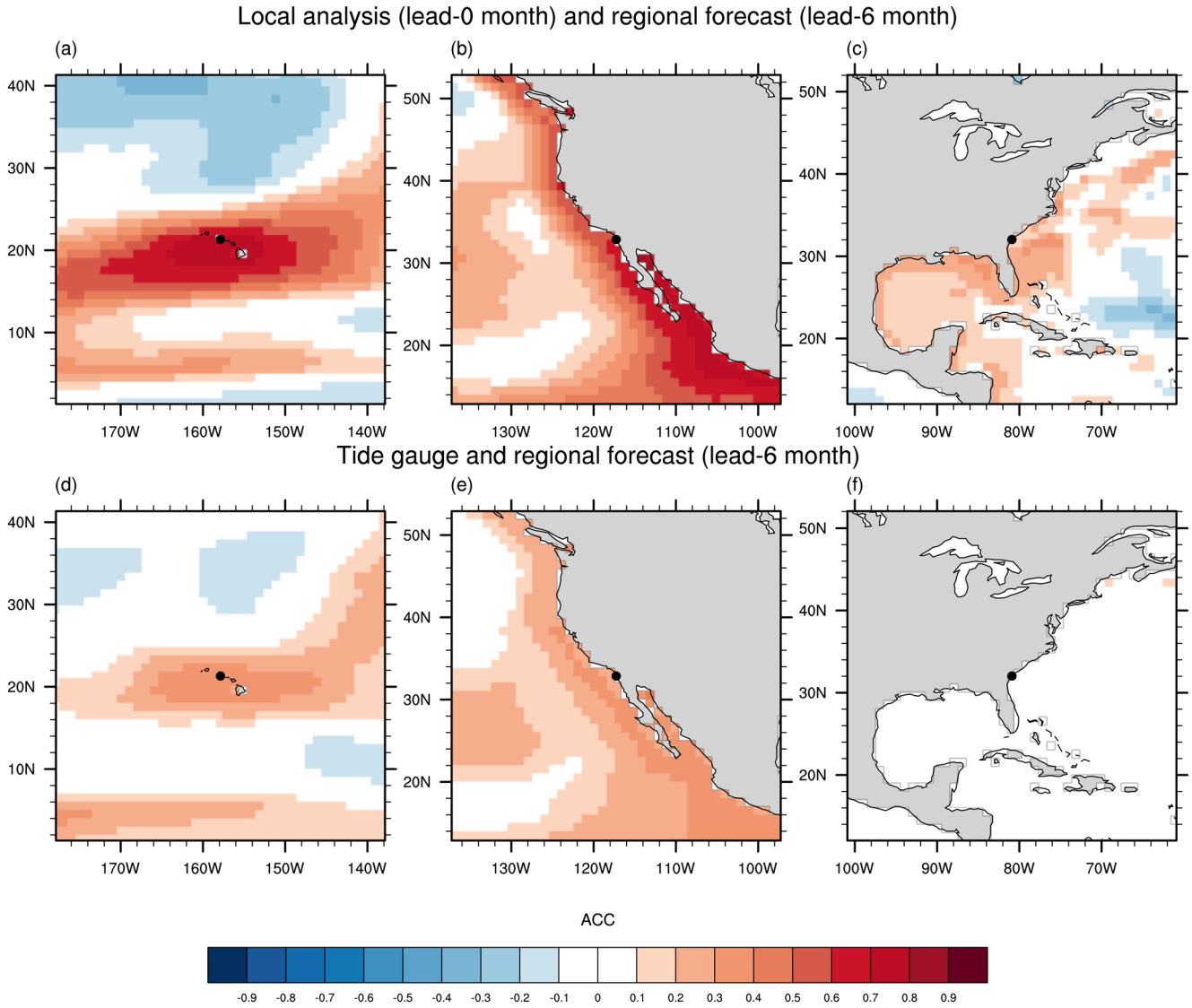


Figure 11. Similar to Figure 10, but considering the spatial correlation patterns associated with lead-6 month forecasts. Two correlation maps are shown for each location: (a–c) local analysis (a single-member forecast at lead-0 month is used to approximate the model analysis) and regional forecast (multi-model mean at the lead-6 months); (d–f) tide gauge and regional forecast at the lead-6 months.

typically too homogeneous and robust, especially around Fort Pulaski and throughout the US Southeast Coast (see also around Puerto Rico; Figure S7). In contrast, the correlation patterns between the tide gauge observations and the lead-0 month forecasts (Figures 10g–10j) are generally weaker than what is either observed or internal to the forecast models. Pattern-wise, whereas the tide gauge-model lead-0 month correlations mostly match the observations around Honolulu and La Jolla (Figures 10g and 10h vs. Figures 10a and 10b), there is almost no evidence of a US East Coast signal around Fort Pulaski (Figure 10i vs. Figure 10c), at least under the imposed constraint of considering how well the multi-model mean variability is in phase with tide gauge observations.

Such coastal discrepancies for the initial forecast month, as measured by the correlation between the tide gauge observation and model during the lead-0 month, can potentially continue throughout the forecast period. Figure 11 shows the point-wise correlation patterns for lead-6 month forecasts at Honolulu, La Jolla, and Fort Pulaski compared to either the model lead-0 month (i.e., approximating the model initial conditions/analysis; panels a–c) or the tide gauge observations (d–f). For Fort Pulaski, the lead-6 month forecast compared to lead-0 month shows a signal along the US Southeast Coast that somewhat resembles

observations (Figure 11c vs. Figure 10c). If instead the lead-6 month forecast is compared to the tide gauge observation, then the correlation pattern is nonexistent (Figure 10f). The latter correlation pattern is illustrative of a stark deficiency in the ability to skillfully forecast sea-level variability in the region; that is, not only at Fort Pulaski, but also along the entire US Gulf of Mexico and Southeast Coasts (see Figure S8 for other locations globally). For Honolulu and La Jolla, the differences are more subtle between lead-6 month correlation patterns using either the lead-0 month (Figures 11a and 11b) or the tide gauge observation (Figures 11d and 11e), but the magnitudes are still much larger for the former, which suggests both challenges and opportunities related to sea-level forecasting in these regions that we will discuss in Section 4.

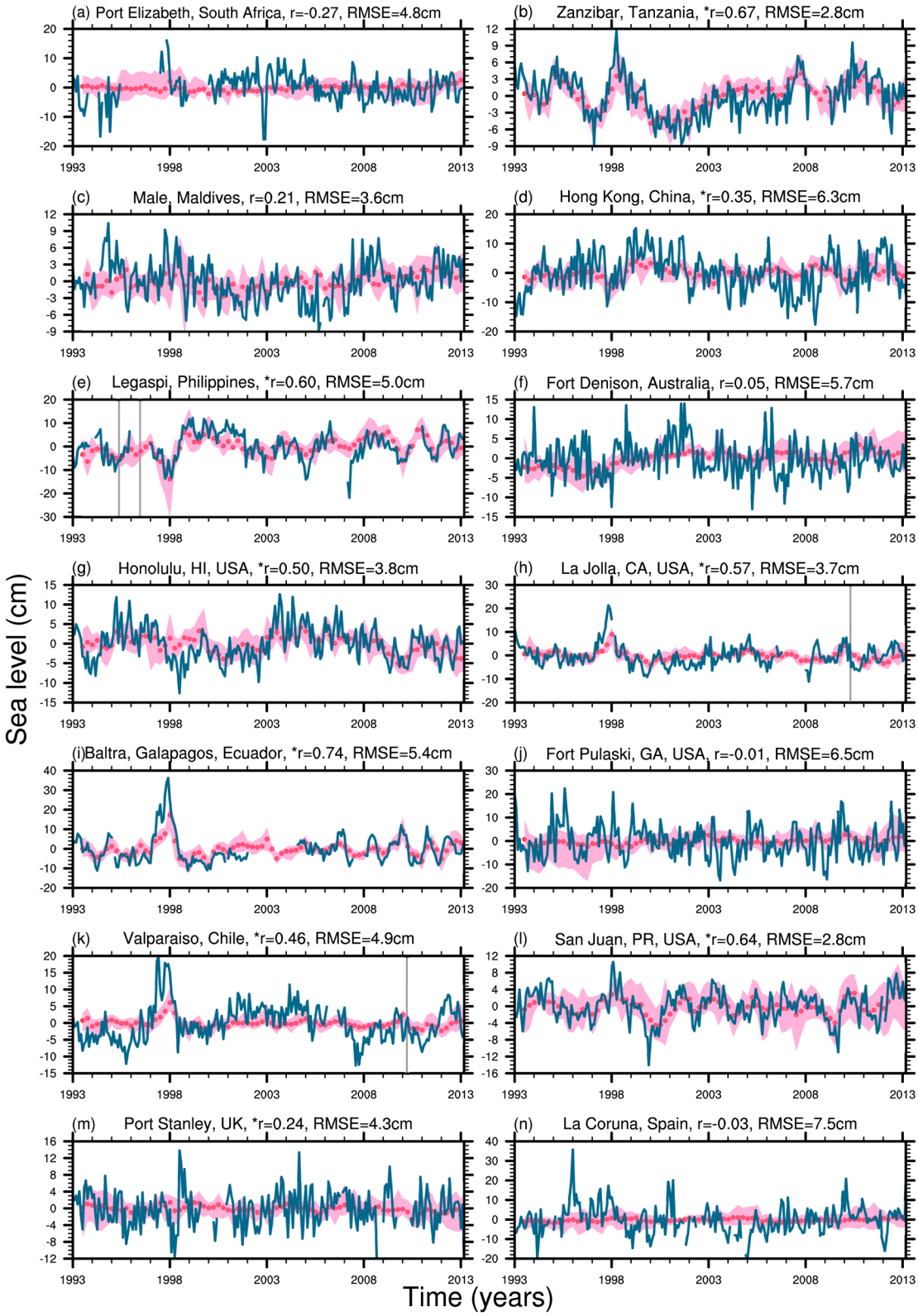
The global forecast assessment identified obvious regional differences in skill (e.g., Figure 8). At the 14 tide gauge locations selected, the multi-model mean forecast at 6 month lead is skillful at only nine places (i.e., statistically significant positive correlations; Figure 12). Six of these nine locations are in the tropics (Zanzibar, Hong Kong, Legaspi, Honolulu, Baltra, and San Juan), where most of the 10 models resolve the observed sea-level variability well (Figure 1). Two other locations with skillful forecasts (La Jolla and Valparaíso) are along the eastern boundary of the higher-latitude Pacific, where ENSO has been shown to most strongly influence sea-level variability outside of the tropics (McIntosh et al., 2015; Menéndez & Woodworth, 2010). The Pacific sea level response to ENSO is well predicted by the multi-model forecasts (e.g., the 1997/1998 El Niño event), not only in the tropics (e.g., at Baltra in the Galapagos Islands; Figure 12i), but also at La Jolla and Valparaíso (Figures 12h and 12k, respectively). For these locations where the lead-6 month forecasts are considered skillful based on the ACC metric, the RMSE values are also smaller than the standard deviations of the tide gauge observations (e.g., at Baltra where the forecast error is 5.4 cm vs. an observed standard deviation of 7.6 cm; comparing Figures 12i and 2i). At the other six locations with poor forecast skill (i.e., either the five with non-significant ACC values, or the marginally significant case of $r = 0.24$ at Port Stanley), the RMSE values tend to be higher relative to the observed variability (e.g., a forecast error of 6.5 cm at Fort Pulaski compared to an observed standard deviation of 7.2 cm; Figures 12j and 2j, respectively).

Considering the ACC metrics for the multi-model mean lead-6 month forecasts shown in Figure 12, there are five locations of no skill according to statistical significance testing that we applied. Poorly performing forecasts are common for locations near the western boundary currents of the Pacific or Atlantic Oceans (i.e., Fort Denison and Fort Pulaski; labeled F and J on the map in Figure 1d) as well as in regions of strong mesoscale eddy activity in the Southern Hemisphere midlatitudes (i.e., Port Elizabeth; A). The two other locations where forecasts performed poorly are at Male and La Coruna (C and N, respectively), which represent deficiencies of the multi-model ensemble in parts of the tropical Indian Ocean and in most of the Atlantic Ocean (see again Figure 8 for global maps of the forecast skill).

4. Discussion

The multi-model assessment identified a number of challenges to the development of skillful forecasts of monthly and seasonal sea-level variability, as well as some opportunities. Most of the challenges in utilizing existing seasonal forecast systems are related to the forecast skill for sea level, which we showed is spatially heterogeneous and usually worse for coastal locations compared to the neighboring open ocean (e.g., comparing the ACC values in Figures 8 and 12). The current poor forecast skill for many of the world's coastal locations is perhaps the most daunting challenge from an applications perspective, since the usefulness of sea-level forecast products will ultimately depend on their ability to accurately predict future variability on monthly-to-seasonal timescales. However, this assessment also provided insights about how climate models forecast sea-level variability, which may aid planning efforts in applying seasonal forecasting science toward development of sea-level forecasting activities.

Design of seasonal sea-level forecast products may begin with assessment of what information is most needed to meet users' needs in a location. With ongoing sea-level rise, coastal stakeholders are increasingly aware of water level offsets between the tidal prediction and what is observed (Sweet et al., 2020). To better anticipate the coastal sea-level conditions, we suggest developing a new type of high-water seasonal alert calendar (e.g., Stephens et al., 2014) that includes outlooks of monthly-to-seasonal sea-level variability, in addition to existing information about long-term sea-level change from observations, as well as tides, which


(a) Port Elizabeth, South Africa, r=-0.27, RMSE=4.8cm
(b) Zanzibar, Tanzania, *r=0.67, RMSE=2.8cm
(c) Male, Maldives, r=0.21, RMSE=3.6cm
(d) Hong Kong, China, *r=0.35, RMSE=6.3cm
(e) Legaspi, Philippines, *r=0.60, RMSE=5.0cm
(f) Fort Denison, Australia, r=0.05, RMSE=5.7cm
(g) Honolulu, HI, USA, *r=0.50, RMSE=3.8cm
(h) La Jolla, CA, USA, *r=0.57, RMSE=3.7cm
(i)Baltra, Galapagos, Ecuador, *r=0.74, RMSE=5.4cm
(j) Fort Pulaski, GA, USA, r=-0.01, RMSE=6.5cm
(k) Valparaiso, Chile, *r=0.46, RMSE=4.9cm
(l) San Juan, PR, USA, *r=0.64, RMSE=2.8cm
(m) Port Stanley, UK, *r=0.24, RMSE=4.3cm
(n) La Coruna, Spain, r=-0.03, RMSE=7.5cm
Sea level (cm)
Time (years)

are typically calculated using harmonic analysis methods and based on known astronomical cycles (Codiga, 2011). Such calendars could be presented to users in a similar construct as existing and widely available tide calendars, but with added information about whether local sea levels (and thus tidal cycles) are likely to be above or below normal, up to several months in advance. Besides the sea-level outlook and tidal prediction, information about the rate of sea-level change would also need to be included so that the sea-level forecast anomaly is referenced to the recently observed sea-level climatology. Although none of the seasonal forecasting models here resolve sea-level changes due to land motion (cf. post-earthquake forecast offsets in Figure 12, consequences of which were noted in Widlansky et al., 2017), and only two of the models that we considered include global sea-level rise (ACCESS-S1 and SEAS5; Figure 3), the long-term trend of sea level is available via calculation from either tide gauges (Figure 2) or altimetry observations (Figure 3a); the former of which is relative to any coastal elevation changes.

As explained in Section 2, we removed any long-term trend from each model at each grid point to achieve an equitable assessment across models of their seasonal forecasting skill. If instead we retained the long-term trends in our assessment (Figure 13), then the lead-6 month ACC improves (worsens) in most places for the two (eight) models that include global sea-level rise (or not). The disparity in forecast skill depending on how the trend is treated is most pronounced in the tropical Atlantic Ocean, and also along the US East Coast, whereas there is less of a difference across models in the tropical Pacific. Since including the trend actually slightly diminishes the skill in the tropical Pacific, it is possible that we have somewhat overestimated the forecast skill there (e.g., if our analyses are biased by whether, or not, particular ENSO events are included in the 20 years of retrospective forecasts). A longer set of seasonal forecasts, extending beyond 2012, would be helpful to better assess how long-term trends affect skill across models.

In terms of developing high-water alert calendars, the ability to skillfully predict coastal sea-level variability at monthly-to-seasonal leads is the most important determinant for achieving a useful product. In the multi-model skill assessment, we used metrics of ACC and RMSE to describe, respectively, the retrospective forecast skill related to prediction of the phase and amplitude of sea-level variability. Unfortunately, the assessment identified many regions and coastal locations where ACC and RMSE values for the multi-model mean forecasts are both poor (i.e., low ACC and high RMSE compared to observed variability; e.g., at La Coruna in the midlatitude North Atlantic, Figure 12n). For many places outside of the tropics, individual models did not perform well either; however, we did notice a subtle advantage in the midlatitudes for the models that assimilate altimetry and also have higher resolution. Furthermore, comparisons of the ACC as a function of lead time between the multi-model mean and individual model forecasts (Figure 9 and Figure S6) did not reveal a systematic tendency for the former to perform better, at least for the majority of sample coastal locations. Throughout most of the tropics, however, we found that the multi-model mean forecast is skillful (e.g., noting the statistically significant ACC at lead-6 months for Zanzibar, Hong Kong, Legaspi, Honolulu, Baltra, and San Juan; Figure 12). Also, we showed that when assessed globally, the multi-model mean lead-6 month forecast performs better than any individual model at the majority of locations (Figures 6a and 7). Overall, we found that useful seasonal outlooks of sea level are achievable in some locations using existing forecast systems and the post-processing methods applied here (i.e., removal of any lead-time and initial-time dependent model biases as well as subtracting long-term trends, which were all done on a location-by-location basis).

Communicating forecast utility and uncertainty is also important for successful implementation of any product. Utility of the multi-model mean forecast is related to the magnitude of the ACC that we calculated (e.g., presented globally in Figure 8 and for the coastal locations in Figure 9 as well as Figure S6). The inter-model forecast spread shown in Figure 12 also includes uncertainty information, which could be incorporated into a real-time product (e.g., by communicating if a particular grouping of models is forecasting a different evolution of the sea level compared to the others). In the simplest form, the skill metrics calculated from the retrospective forecasts could be used to mask regions of the world where useful sea-level outlooks

Figure 12. Observed and forecasted sea-level monthly anomalies (blue and red, respectively) during 1993–2012 (2013 predictions) for the 14 sample locations. Retrospective lead-6 month forecasts are started four times per year (January, April, July, and October). Anomaly correlation coefficient (r, * indicate if the correlation is significantly different from 0) and root-mean-square error (RMSE) between tide gauge observations and multi-model mean forecasts (dots) are noted. Shading indicates the inter-model spread, which is interpolated between forecast target times (July, October, January, and April). As in Figure 2, times of earthquakes occurring within 200 km of a location and exceeding 7.5Mw are indicated by vertical lines, which may indicate potential vertical land motion.

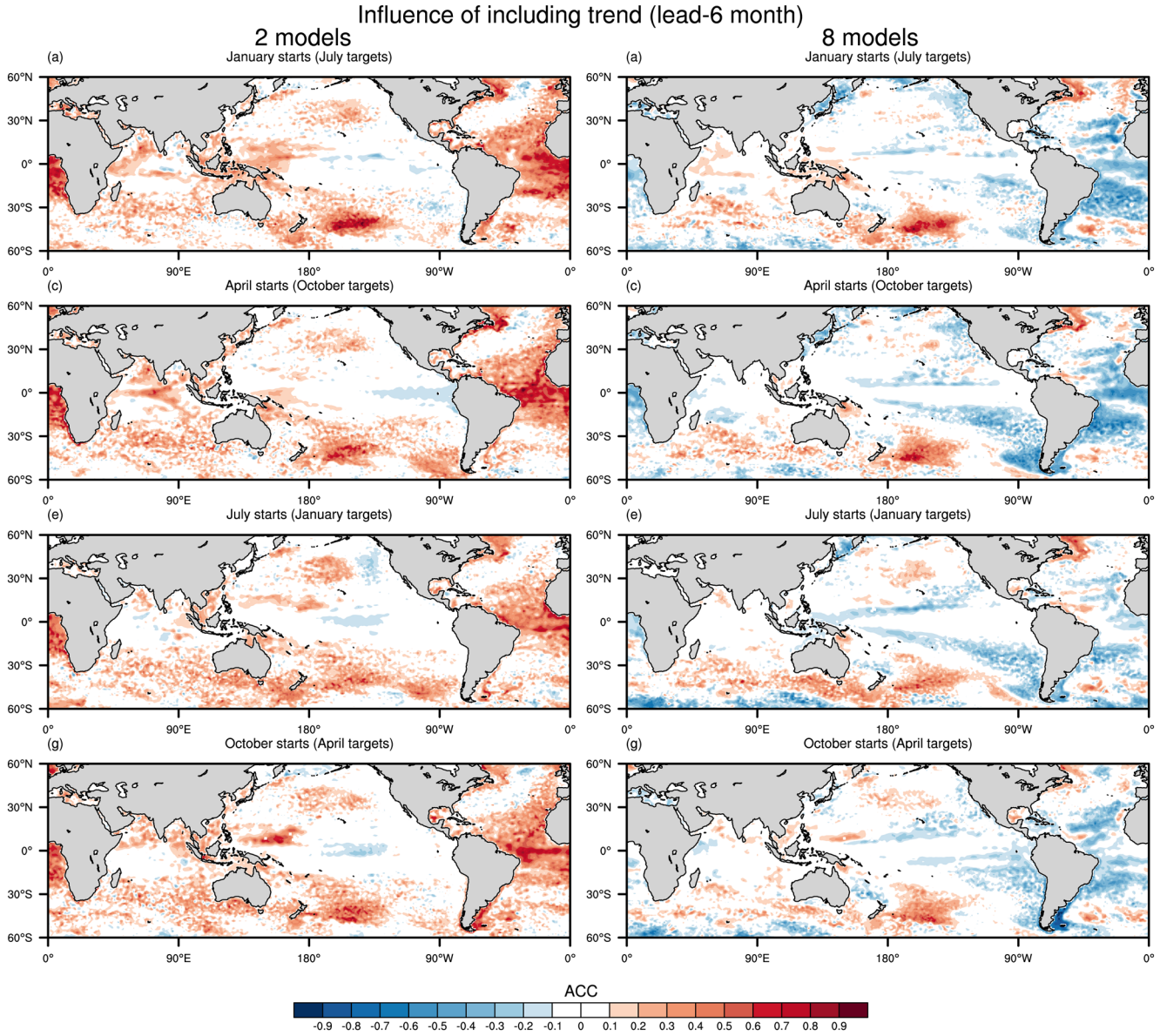


Figure 13. Influence of including the trend on retrospective forecast skills (anomaly correlation coefficient [ACC]) at lead-6 months for forecasts of monthly sea-level anomalies. Shown are the ACC differences between verifications using either no detrending of forecasts and observations (trends included), or when the linear trend is removed from both (as we did for all the other analyses). The left column is the average of two models that include global sea-level rise (ACCESS-S1 and SEAS5) and the right column is the average of the other eight models. Forecasts are initialized during January (a), April (b), July (c), or October (d) and compared to altimetry observations during the following July, October, January (year +1), or April (year +1), respectively.

are not yet practical (e.g., by asking whether the ACC is either statistically significant or at least improved compared to a simple damped-persistence forecast). Furthermore, there are places where the observed variability of monthly sea-level anomalies is so small (e.g., the standard deviation in parts of the equatorial Atlantic is less than 2 cm; Figure 1a) that a seasonal forecast is perhaps unnecessary. We note also that forecast skill in the tropical Atlantic is clearly worse than at similar latitudes of the Pacific and Indian Oceans (Figure 8), which is not surprising considering that we would expect any sea-level anomaly in the former region to be small in both the forecasts and observations (i.e., a low signal-to-noise ratio is anticipated in most of the tropical Atlantic).

Whereas the multi-model mean forecast usually outperforms any individual model globally (Figures 6a and 7), lack of clear evidence of such an advantage at the coastal locations that we examined suggests that

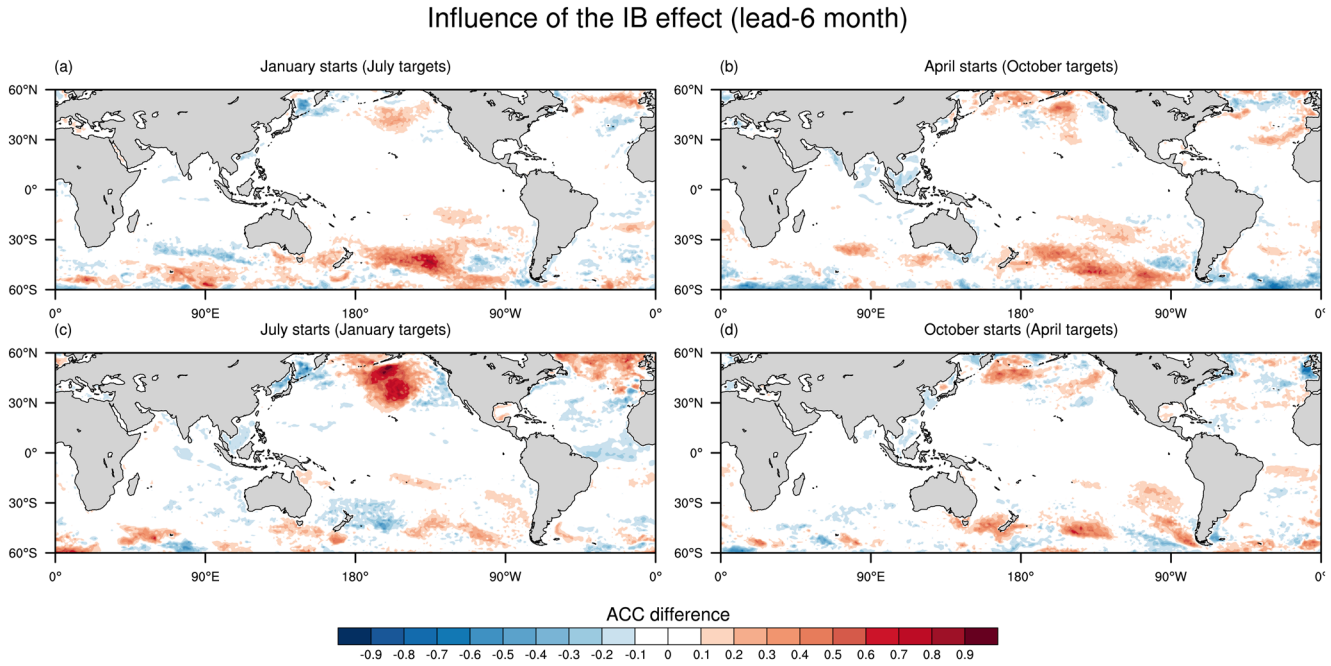


Figure 14. Influence of the inverse-barometer (IB) effect on retrospective forecast skills (anomaly correlation coefficient [ACC]) at lead-6 months for the multi-model mean forecasts of monthly sea-level anomalies. Shown are the ACC differences between verification using altimetry without the IB effect minus with the IB effect included. Forecasts, which do not include the IB effect, are initialized during January (a), April (b), July (c), or October (d) and compared to altimetry observations during the following July, October, January (year +1), or April (year +1), respectively.

unequal performance across the 10 models (Figure 5 and Figure S4) may be inhibiting the potential to skillfully predict monthly sea-level anomalies. Forecast improvements could potentially come from modeling advancements such as using higher spatial resolution ocean models to better resolve coastal variability. Another opportunity for forecasting improvement is through better ocean initialization, the importance of which is suggested by the linear dependence of the forecast skill at lead-6 months versus how well the sea-level anomaly forecast at lead-0 month resembles the observation (Figure 6). In the absence of any improvements to the models or assimilation systems, it is also possible that using skill-based weighting approaches to design a more sophisticated multi-model ensemble may yield more skillful forecasts; however, considering the lack of any model to forecast well the sea-level variability at some of the coastal locations we considered (e.g., at Fort Pulaski on the US East Coast; Figure 12j), we are disinclined to pursue this approach.

Inclusion of the IB effect in the forecast, either through modeling advancements or by adding to the post-processing routine the atmospheric pressure forcing term (Yin et al., 2020), may also potentially improve the skill, especially in the midlatitudes. To estimate how much of a forecasting improvement may come from including atmospheric pressure forcing, we show in Figure 14 the multi-model mean ACC if the lead-6 month forecast is verified using altimetry with the IB effect removed (i.e., using the Dynamic Atmospheric Correction that is commonly applied to altimetry products), compared to our original diagnostic (see Figure 8). Figure S9 shows a similar comparison, but for the lead-0 month. We note that positive ACC values in these comparisons represent a deficiency of sea-level forecasting, since none of the current-generation models include the IB effect. In the midlatitudes, the ACC differences are noticeable, especially during winters of the respective hemispheres (Figures 14a and 14c). There are also some places where including the IB effect is not likely to improve the 6-month forecast of sea level, such as along the US East and West Coasts, which is not surprising considering the challenges of predicting atmospheric pressure on seasonal timescales. Since including the IB effect would be unlikely to diminish much the forecast skill anywhere at short leads (e.g., lead-0 month; Figure S9), and would likely improve the forecast in parts of the midlatitudes at longer leads (e.g., lead-6 month; Figure 14), this is a relatively simple way forward to achieving higher skills in many places.

We are encouraged that our assessment of the individual models highlighted clear differences in forecast skill over large regions (e.g., differences in global patterns of both the lead-0 and lead-6 month ACC between the CFSv2, CFSv2-COLA, and SEAS5 models; Figure 5). The inter-model assessment can potentially be used to inform pathways toward effectively improving next-generation seasonal forecast systems, as it relates to making better predictions of monthly sea-level anomalies. Notably for CFSv2 and CFSv2-COLA (Figure 5), improving the ocean assimilation in the latter forecasting system while otherwise having the same model, clearly improved the lead-0 month sea-level anomaly compared to observations, and also somewhat improved the lead-6 month forecast. Interestingly, when comparing the lead-6 month forecasts instead to the sea-level variability internal to the respective forecasting systems (i.e., using lead-0 month forecasts; e.g., Figures 5g–5i), the ACC values are significant nearly everywhere and the inter-model differences in skill are overall more subtle. If assessing forecast skill this way, in some places the CFSv2 model appears to perform better than the models that are known to have improved assimilation systems such as CFSv2-COLA and SEAS5 (as measured by their comparisons with altimetry observations during the lead-0 month; Figures 5a–5c). The substantially higher forecast skill in regards to models predicting their own sea-level anomalies (i.e., variability during the lead-0 month; see also Figure 11 and Figure S8) may perhaps be considered an opportunity in that the result suggests widespread predictability of sea-level variability. We caution, however, that since the real forecast skill (as measured by comparison to observed sea-level anomalies) more closely resembles how a sea-level outlook will be verified in practice, this should be the primary assessment of model performance.

Overall the two forecasting systems that directly assimilate information from altimetry observations of SSH in their initial conditions, SEAS5 (Figure 5) and ACCESS-S1 (Figure S4), tend to perform better than any of the other eight models. However, considering that there are also other differences between the models (e.g., higher spatial resolution of the ocean models), we are not able to conclusively identify the use of altimetry measurements in the ocean assimilation as the cause of improved sea-level forecasts. Targeted ocean assimilation experiments that either include, or not, observations of sea level would help better guide the development of new seasonal forecasting systems.

5. Conclusion

The societal need for future sea-level outlooks is increasing with continuing sea-level rise that is already contributing to recurrent high-tide flooding events (Sweet et al., 2020). Sea-level variability, which itself may increase in a warming climate (Widlansky et al., 2020), will continue to cause the clustering in both space and time of coastal high-tide impacts. Positive sea-level anomalies, which are not currently considered in most tidal predictions, may contribute to coastal impacts including flooding, erosion, or saltwater inundation. There is effort currently underway to incorporate seasonal forecasts of sea-level anomalies into existing high-tide outlooks; however, such activity is stymied by an overall lack of sea-level forecast data publicly available compared to other ocean variables such as sea surface temperature.

Real-time forecast output of the SSH variable is only publicly available from one of the 10 seasonal forecasting models that we considered (CFSv2). Other models, such as ACCESS-S1 and SEAS5, do output the SSH as part of their operational forecasting activities, although data use is more restricted. Yet, nearly all of the models that we considered, as well as many others, provide publicly real-time forecasts of other oceanic variables (primarily sea surface temperature) via collaborative frameworks such as the Subseasonal-to-Seasonal (S2S) Prediction Project established by the World Climate Research Program (Vitart et al., 2017). Despite SSH being identified as a parameter of interest by S2S, few organizations currently contribute real-time output of this variable.

For the locations that we have identified as having skillful seasonal sea-level forecasts, developing useful outlooks of future high-water events may be as simple as combining the SSH output from existing climate forecast systems with standard tide predictions, which is already being done for many tropical Pacific islands (Widlansky et al., 2017) but not elsewhere. In other places where current-generation forecasting systems cannot skillfully predict sea-level variability, technological improvements are required in the models, their initialization schemes, and/or post-processing approaches. All of these activities would benefit from

more widely available output of seasonal sea-level forecasting data to support further skill assessments as well as making real-time predictions.

Data Availability Statement

The data used in this study are available from the following sources: tide gauge observations (http://uhslc.soest.hawaii.edu/data/?rq), altimetry observations (https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=SEALEVEL_GLO_PHY_L4_REP_OBSERVATIONS_008_047), and retrospective forecasts (http://uhslc.soest.hawaii.edu/opendap/NMMEhindcasts/contents.html).

Acknowledgments
This study was supported by the NOAA Climate Program Office's Modeling, Analysis, Predictions, and Projections (MAPP) program. The authors acknowledge the helpful collaborations facilitated by the MAPP-organized Marine Prediction Task Force as well as insightful comments from two anonymous reviewers. The authors also thank the forecasting organizations for producing and making available their model output. Xiaoyu Long, Matthew J. Widlansky, and H. Annamalai were supported by NOAA Grant NA17OAR4310110. Yoshimitsu Chikamoto was supported by the Utah Agricultural Experiment Station, Utah State University (approved as journal paper 9457), SERDP Award RC19-F1-1389, and the US Department of Interior, Bureau of Reclamation (R19AP00149). The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect those of NOAA or the Department of Commerce.

References

Allen, J. S. (1975). Coastal trapped waves in a stratified ocean. *Journal of Physical Oceanography*, *5*(2), 300–325. https://doi.org/10.1175/1520-0485(1975)005<0300:CTWIAS>2.0.CO;2

Anderson, J., Milliken, K., Wallace, D., Rodriguez, A., & Simms, A. (2010). Coastal impact underestimated from rapid sea level rise. *Eos, Transactions, American Geophysical Union*, *91*(23), 205–206. https://doi.org/10.1029/2010EO230001

Balmaseda, M. A., Davey, M. K., & Anderson, D. L. T. (1995). Decadal and seasonal dependence of ENSO prediction skill. *Journal of Climate*, *8*(11), 2705–2715. https://doi.org/10.1175/1520-0442(1995)008⟨2705:DASDOE⟩2.0.CO;2

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1132–1161. https://doi.org/10.1002/qj.2063

Becker, J. M., Merrifield, M. A., & Ford, M. (2014). Water level effects on breaking wave setup for Pacific Island fringing reefs. *Journal of Geophysical Research: Oceans*, *119*(2), 914–932. https://doi.org/10.1002/2013JC009373

Bulgin, C. E., Merchant, C. J., & Ferreira, D. (2020). Tendencies, variability and persistence of sea surface temperature anomalies. *Scientific Reports*, *10*(1), 7986. https://doi.org/10.1038/s41598-020-64785-9

Calafat, F. M., Wahl, T., Lindsten, F., Williams, J., & Frajka-Williams, E. (2018). Coherent modulation of the sea-level annual cycle in the United States by Atlantic Rossby waves. *Nature Communications*, *9*(1), 1–13. https://doi.org/10.1038/s41467-018-04898-y

Caldwell, P., Merrifield, M., & Thompson, P. (2015). Sea level measured by tide gauges from global oceans as part of the Joint Archive for Sea Level (JASL) since 1846 (Vol. 10). National Oceanographic Data Center, NOAA. https://doi.org/10.7289/v5v40s7w

Cazenave, A., & Cozannet, G. L. (2014). Sea level rise and its coastal impacts. *Earth's Future*, *2*(2), 15–34. https://doi.org/10.1002/2013EF000188

Chen, H.-C., & Jin, F.-F. (2020). Fundamental behavior of ENSO phase locking. *Journal of Climate*, *33*(5), 1953–1968. https://doi.org/10.1175/JCLI-D-19-0264.1

Chikamoto, Y., Timmermann, A., Widlansky, M. J., Zhang, S., & Balmaseda, M. A. (2019). A drift-free decadal climate prediction system for the Community Earth System Model. *Journal of Climate*, *32*(18), 5967–5995. https://doi.org/10.1175/JCLI-D-18-0788.1

Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., & others (2013). Sea level change. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, et al. (Eds.), *Climate change 2013: The physical science basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Church, J. A., & White, N. J. (2006). A 20th century acceleration in global sea-level rise. *Geophysical Research Letters*, *33*(1), a–n. https://doi.org/10.1029/2005GL024826

Codiga, D. (2011). *Unified tidal analysis and prediction using the UTide Matlab functions* (Technical Report 2011-01, Graduate School of Oceanography). University of Rhode Island.

Faghmous, J. H., Frenger, I., Yao, Y., Warmka, R., Lindell, A., & Kumar, V. (2015). A daily global mesoscale ocean eddy dataset from satellite altimetry. *Scientific Data*, *2*(1), 150028. https://doi.org/10.1038/sdata.2015.28

Fasullo, J. T., & Nerem, R. S. (2018). Altimeter-era emergence of the patterns of forced sea-level rise in climate models and implications for the future. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(51), 12944–12949. https://doi.org/10.1073/pnas.1813233115

Fraser, R., Palmer, M., Roberts, C., Wilson, C., Copsey, D., & Zanna, L. (2019). Investigating the predictability of North Atlantic sea surface height. *Climate Dynamics*, *53*(3), 2175–2195. https://doi.org/10.1007/s00382-019-04814-0

Fu, L.-L., & Pihos, G. (1994). Determining the response of sea level to atmospheric pressure forcing using TOPEX/POSEIDON data. *Journal of Geophysical Research*, *99*(C12), 24633–24642. https://doi.org/10.1029/94jc01647

Gaspar, P., & Ponte, R. M. (1997). Relation between sea level and barometric pressure determined from altimeter data and model simulations. *Journal of Geophysical Research*, *102*(C1), 961–971. https://doi.org/10.1029/96JC02920

Graham, R., Yun, W., Kim, J., Kumar, A., Jones, D., Bettio, L., et al. (2011). Long-range forecasting and the Global Framework for Climate Services. *Climate Research*, *47*(1–2), 47–55. https://doi.org/10.3354/cr00963

Hamlington, B. D., Leben, R. R., Kim, K. Y., Nerem, R. S., Atkinson, L. P., & Thompson, P. R. (2015). The effect of the El Niño-Southern Oscillation on U.S. regional and coastal sea level. *Journal of Geophysical Research: Oceans*, *120*(6), 3970–3986. https://doi.org/10.1002/2014JC010602

Holbrook, N. J., Claar, D. C., Hobday, A. J., McInnes, K. L., Oliver, E. C. J., Gupta, A. S., et al. (2020). ENSO-driven ocean extremes and their ecosystem impacts. In M. J. McPhaden, A. Santoso, & W. Cai (Eds.), *El Niño southern oscillation in a changing climate* (pp. 409–428). Wiley Online Library. https://doi.org/10.1002/9781119548164.ch18

Huang, B., Shin, C.-S., Shukla, J., Marx, L., Balmaseda, M. A., Halder, S., et al. (2017). Reforecasting the ENSO Events in the Past 57 Years (1958-2014). *Journal of Climate*, *30*(19), 7669–7693. https://doi.org/10.1175/JCLI-D-16-0642.1

Hudson, D., Alves, O., Hendon, H. H., Lim, E.-P., Liu, G., Luo, J.-J., & others (2017). ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system. *Journal of Southern Hemisphere Earth Systems Science*, *67*(3), 132–159. https://doi.org/10.22499/3.6703.001

Hu, D., Wu, L., Cai, W., Gupta, A. S., Ganachaud, A., Qiu, B., et al. (2015). Pacific western boundary currents and their roles in climate. *Nature*, *522*(7556), 299–308. https://doi.org/10.1038/nature14504

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, *12*(3), 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2014). The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*(4), 585–601. https://doi.org/10.1175/BAMS-D-12-00050.1

Krishnamurti, T. N., Chakraborty, A., Krishnamurti, R., Dewar, W. K., & Clayson, C. A. (2006). Seasonal prediction of sea surface temperature anomalies using a suite of 13 coupled atmosphere-ocean models. *Journal of Climate*, *19*(23), 6069–6088. https://doi.org/10.1175/JCLI3938.1. Retrieved from https://journals.ametsoc.org/view/journals/clim/19/23/jcli3938.1.xml

Kumar, A., Hu, Z.-Z., Jha, B., & Peng, P. (2017). Estimating ENSO predictability based on multi-model hindcasts. *Climate Dynamics*, *48*(1–2), 39–51. https://doi.org/10.1007/s00382-016-3060-4

Latif, M., Anderson, D., Barnett, T., Cane, M., Kleeman, R., Leetmaa, A., et al. (1998). A review of the predictability and prediction of ENSO. *Journal of Geophysical Research*, *103*(C7), 14375–14393. https://doi.org/10.1029/97JC03413

Liu, G., Eakin, C. M., Chen, M., Kumar, A., De La Cour, J. L., Heron, S. F., et al. (2018). Predicting heat stress to inform reef management: NOAA coral reef watch's 4-month coral bleaching outlook. *Frontiers in Marine Science*, *5*, 57. https://doi.org/10.3389/fmars.2018.00057

Long, X., Widlansky, M. J., Schloesser, F., Thompson, P. R., Annamalai, H., Merrifield, M. A., & Yoon, H. (2020). Higher sea levels at Hawaii caused by strong El Niño and weak trade winds. *Journal of Climate*, *33*(8), 3037–3059. https://doi.org/10.1175/JCLI-D-19-0221.1

McIntosh, P. C., Church, J. A., Miles, E. R., Ridgway, K., & Spillman, C. M. (2015). Seasonal coastal sea level prediction using a dynamical model. *Geophysical Research Letters*, *42*(16), 6747–6753. https://doi.org/10.1002/2015GL065091

Menéndez, M., & Woodworth, P. L. (2010). Changes in extreme high water levels based on a quasi-global tide-gauge data set. *Journal of Geophysical Research*, *115*(C10). https://doi.org/10.1029/2009JC005997

Merrifield, M. A., Genz, A. S., Kontoes, C. P., & Marra, J. J. (2013). Annual maximum water levels from tide gauges: Contributing factors and geographic patterns. *Journal of Geophysical Research: Oceans*, *118*(5), 2535–2546. https://doi.org/10.1002/jgrc.20173

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., et al. (2013). The Canadian seasonal to interannual prediction system. Part I: Models and initialization. *Monthly Weather Review*, *141*(8), 2910–2945. https://doi.org/10.1175/MWR-D-12-00216.1

Miles, E. R., Spillman, C. M., Church, J. A., & McIntosh, P. C. (2014). Seasonal prediction of global sea level anomalies using an ocean-atmosphere dynamical model. *Climate Dynamics*, *43*(7–8), 2131–2145. https://doi.org/10.1007/s00382-013-2039-7

Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., Feldman, D. L., Sweet, W., Matthew, R. A., & Luke, A. (2015). Increased nuisance flooding along the coasts of the United States due to sea level rise: Past and future. *Geophysical Research Letters*, *42*(22), 9846–9852. https://doi.org/10.1002/2015GL066072

Nicholls, R. J., & Cazenave, A. (2010). Sea-level rise and its impact on coastal zones. *Science*, *328*(5985), 1517–1520. https://doi.org/10.1126/science.1185782

Palanisamy, H., Cazenave, A., Meyssignac, B., Soudarin, L., Wöppelmann, G., & Becker, M. (2014). Regional sea level variability, total relative sea level rise and its impacts on islands and coastal zones of Indian Ocean over the last sixty years. *Global and Planetary Change*, *116*, 54–67. https://doi.org/10.1016/j.gloplacha.2014.02.001

Penduff, T., Juza, M., Brodeau, L., Smith, G. C., Barnier, B., Molines, J.-M., et al. (2010). Impact of global ocean model resolution on sea-level variability with emphasis on interannual time scales. *Ocean Science*, *6*, 269–284. https://doi.org/10.5194/os-6-269-2010

Piecuch, C. G., Dangendorf, S., Ponte, R. M., & Marcos, M. (2016). Annual sea level changes on the North American Northeast Coast: Influence of local winds and barotropic motions. *Journal of Climate*, *29*(13), 4801–4816. https://doi.org/10.1175/JCLI-D-16-0048.1

Roberts, C. D., Calvert, D., Dunstone, N., Hermanson, L., Palmer, M. D., & Smith, D. (2016). On the drivers and predictability of seasonal-to-interannual variations in regional sea level. *Journal of Climate*, *29*(21), 7565–7585. https://doi.org/10.1175/JCLI-D-15-0886.1

Román-Rivera, M. A., & Ellis, J. T. (2018). The king tide conundrum. *Journal of Coastal Research*, *344*(4), 769–771. https://doi.org/10.2112/JCOASTRES-D-18A-00001.1

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, *27*(6), 2185–2208. https://doi.org/10.1175/10.1175/JCLI-D-12-00823.1

Shin, S. I., & Newman, M. (2021). Seasonal predictability of global and North American coastal sea surface temperature and height anomalies. *Geophysical Research Letters*, *48*. https://doi.org/10.1029/2020GL091886

Smith, D. M., Eade, R., & Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, *41*(11–12), 3325–3338. https://doi.org/10.1007/s00382-013-1683-2

Stephens, S. A., Bell, R. G., Ramsay, D., & Goodhue, N. (2014). High-water alerts from coinciding high astronomical tide and high mean sea level anomaly in the Pacific Islands region. *Journal of Atmospheric and Oceanic Technology*, *31*(12), 2829–2843. https://doi.org/10.1175/JTECH-D-14-00027.1

Sweet, W., Dusek, G., Carbin, G., Marra, J. J., Marcy, D. C., & Simon, S. (2020). *2019 state of US high tide flooding with a 2020 outlook* (Technical Report). National Oceanic and Atmospheric Administration.

Sweet, W. V., & Park, J. (2014). From the extreme to the mean: Acceleration and tipping points of coastal inundation from sea level rise. *Earth's Future*, *2*(12), 579–600. https://doi.org/10.1002/2014EF000272

Sweet, W. V., & Zervas, C. (2011). Cool-season sea level anomalies and storm surges along the U.S. East Coast: Climatology and comparison with the 2009/10 El Niño. *Monthly Weather Review*, *139*(7), 2290–2299. https://doi.org/10.1175/MWR-D-10-05043.1

Thompson, P. R., Widlansky, M. J., Merrifield, M. A., Becker, J. M., & Marra, J. J. (2019). A statistical model for frequency of coastal flooding in Honolulu, Hawaii, during the 21st century. *Journal of Geophysical Research: Oceans*, *124*(4), 2787–2802. https://doi.org/10.1029/2018JC014741

Torrence, C., & Webster, P. J. (1998). The annual cycle of persistence in the El Nño/Southern Oscillation. *Quarterly Journal of the Royal Meteorological Society*, *124*(550), 1985–2004. https://doi.org/10.1002/qj.49712455010

Tribbia, J. (2015). *NCAR contribution to a U.S. National Multi-Model Ensemble (NMME) ISI prediction system* (Technical Report). University Corporation for Atmospheric Research. https://doi.org/10.2172/1226920

Vannitsem, S., Wilks, D. S., & Messner, J. (2018). *Statistical postprocessing of ensemble forecasts*. Elsevier.

Vinogradov, S. V., & Ponte, R. M. (2010). Annual cycle in coastal sea level from tide gauges and altimetry. *Journal of Geophysical Research*, *115*(C4). https://doi.org/10.1029/2009JC005767

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, *98*(1), 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1

Webster, P. J., Moore, A. M., Loschnigg, J. P., & Leben, R. R. (1999). Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997-98. *Nature*, *401*(6751), 356–360. https://doi.org/10.1038/43848

Widlansky, M. J., Long, X., & Schloesser, F. (2020). Increase in sea level variability with ocean warming associated with the nonlinear thermal expansion of seawater. *Communications Earth & Environment*, *1*(1), 9. https://doi.org/10.1038/s43247-020-0008-8

Widlansky, M. J., Marra, J. J., Chowdhury, M. R., Stephens, S. A., Miles, E. R., Fauchereau, N., et al. (2017). Multimodel ensemble sea level forecasts for tropical Pacific islands. *Journal of Applied Meteorology and Climatology*, *56*(4), 849–862. https://doi.org/10.1175/JAMC-D-16-0284.1

Widlansky, M. J., Timmermann, A., & Cai, W. (2015). Future extreme sea level seesaws in the tropical Pacific. *Science Advances*, *1*(8), e1500560. https://doi.org/10.1126/sciadv.1500560

Widlansky, M. J., Timmermann, A., McGregor, S., Stuecker, M. F., & Cai, W. (2014). An interhemispheric tropical sea level seesaw due to El Niño Taimasa. *Journal of Climate*, *27*(3), 1070–1081. https://doi.org/10.1175/JCLI-D-13-00276.1

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100). Academic Press.

Wöppelmann, G., & Marcos, M. (2016). Vertical land motion as a key to understanding sea level change and variability. *Reviews of Geophysics*, *54*(1), 64–92. https://doi.org/10.1002/2015RG000502

Xue, Y., Chen, M., Kumar, A., Hu, Z.-Z., & Wang, W. (2013). Prediction skill and bias of tropical Pacific sea surface temperatures in the NCEP Climate Forecast System version 2. *Journal of Climate*, *26*(15), 5358–5378. https://doi.org/10.1175/JCLI-D-12-00600.1

Yin, J. (2012). Century to multi-century sea level rise projections from CMIP5 models. *Geophysical Research Letters*, *39*(17), a–n. https://doi.org/10.1029/2012GL052947

Yin, J., Griffies, S. M., Winton, M., Zhao, M., & Zanna, L. (2020). Response of Storm-Related Extreme Sea Level along the U.S. Atlantic Coast to Combined Weather and Climate Forcing. *Journal of Climate*, *33*(9), 3745–3769. https://doi.org/10.1175/jcli-d-19-0551.1

Zhang, L., Han, W., Li, Y., & Lovenduski, N. S. (2019). Variability of sea level and upper-ocean heat content in the Indian Ocean: Effects of subtropical Indian Ocean Dipole and ENSO. *Journal of Climate*, *32*(21), 7227–7245. https://doi.org/10.1175/JCLI-D-19-0167.1

Zhang, S., Harrison, M. J., Rosati, A., & Wittenberg, A. (2007). System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, *135*(10), 3541–3564. https://doi.org/10.1175/MWR3466.1